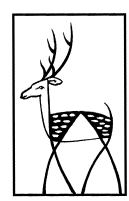
RECRAFTING THE RULE OF LAW

Recrafting the Rule of Law: The Limits of Legal Order

Edited by
DAVID DYZENHAUS
The University of Toronto



OXFORD – PORTLAND OREGON
1999

Hart Publishing
Oxford and Portland, Oregon

Published in North America (US and Canada) by
Hart Publishing c/o
International Specialized Book Services
5804 NE Hassalo Street
Portland, Oregon
97213-3644
USA

Distributed in the Netherlands, Belgium and Luxembourg by
Intersentia, Churchillaan 108
B2900 Schoten
Antwerpen
Belgium

Distributed in Australia and New Zealand by
Federation Press
John St
Leichhardt
NSW 2000

Hart Publishing Ltd is a specialist legal publisher based in Oxford, England.
To order further copies of this book or to request a list of other
publications please write to:

Hart Publishing Ltd, 19 Whitehouse Road, Oxford, OX1 4PA
Telephone: +44 (0)1865 434459 or Fax: +44 (0)1865 794882
e-mail: mail@hartpub.co.uk

British Library Cataloguing in Publication Data
Data Available
ISBN 1 901362–31–0 (cloth)

Typeset by Hope Services (Abingdon) Ltd.
Printed in Great Britain on acid-free paper
by Biddles Ltd, Guildford and Kings Lynn.

Contents

List of Contributors

Richard L. Abel is Connell Professor of Law at UCLA.

Robert Alexy is Professor of Public Law and Legal Philosophy at the University of Kiel.

David Dyzenhaus is Professor of Law and Philosophy at the University of Toronto.

Anton Fagan is Senior Lecturer in Law at the University of Cape Town.

Jody Freeman is Acting Professor of Law at UCLA.

Alon Harel is Senior Lecturer in Law at the Hebrew University, Jerusalem.

Allan C. Hutchinson is Professor of Law at Osgoode Hall Law School.

Neil MacCormick is Regius Professor of Public Law and Leverhulme Personal Research Professor at the University of Edinburgh.

John P. McCormick is Assistant Professor of Political Science at Yale University.

Henry S. Richardson is Professor of Philosophy at Georgetown University.

Julian Rivers is Lecturer in Law at the University of Bristol.

Andras Sajo is Professor of Constitutional Law at the Central European University, Budapest.

William E. Scheuerman is Associate Professor of Political Science at the University of Pittsburgh.

Anthony J. Sebok is Professor of Law at Brooklyn Law School.

Christine Sypnowich is Professor of Philosophy at Queen's University, Kingston.

Kenneth Winston is Lecturer in Ethics and Assistant to the Dean for Special Projects on International Ethics and Diversity at the Kennedy School of Government, Harvard University.

Preface and Acknowledgements

This collection of essays on the rule of law focuses on the traditional question whether the rule of law is necessarily the rule of moral principles, the question of the legitimacy of law. Essays by lawyers, philosophers, and political theorists illuminate and take forward both that question and debate about issues to do with the reach of the rule of law which complicate its answer.

The essays divide into three main sections. The first contains studies of legal orders where the rule of law is under severe stress. These essays are united by their focus on the question of what one might conclude from such a study about the idea of the rule of law. The second section contains essays which engage the question of the value of the rule of law as a conceptual problem within philosophy of law, political philosophy, and legal theory. The third section contains essays which focus on the rule of law as a question about the limits of legal order. That is, they ask whether the rule of law has any application in areas where it might be thought to be only minimally effective—in transitions from dictatorship to democracy, in the control of the process of globalisation, and in the control of administrative agencies and executive discretion.

In putting together this collection I have enjoyed the support of my academic heads of department at the University of Toronto, Ron Daniels (Dean of Law) and Mark Thornton (Chair of Philosophy), as well as generous funding from the Social Sciences and Humanities Research Council of Canada. It gives me great pleasure to record again my debt to Richard Hart, a dedicated publisher.

1

Recrafting the Rule of Law

DAVID DYZENHAUS

INTRODUCTION

The rule of law is often claimed to be one of the ingredients of legitimate government. But one might want to be suspicious of this claim when one notes that it is usually governments who say that their rule is legitimate. In a customarily acerbic and penetrating essay, Judith Shklar called such a claim "just another of those self-congratulatory rhetorical devices that grace the utterances of Anglo–American politicians". "No intellectual effort", she said, "need therefore be wasted on this bit of ruling-class chatter".[1]

Shklar did not, however, conclude that attention to the idea of the rule of law is a waste of effort. In her own survey of the history of the rule of law, she sketched two models, one (associated with Aristotle) which presents the rule of law as the "rule of reason", the other (associated with Montesquieu) which "sees the rule of law as those institutional restraints that prevent governmental agents from oppressing the rest of society".[2] Shklar quite clearly preferred the more modest institutional restraint model as she criticised this century's most forceful exponent of the rule of reason model, Ronald Dworkin, for both naivety and parochialism.

These criticisms attached to Dworkin's famous metaphor of Judge Hercules, the embodiment of the rule of reason, who is able in any particular case of challenge to the law to find a moral justification for the law in a set of principles which justify the law of that legal order, even law and legal order in general. The claim that the rule of law is the rule not only of reason or principles but of sound moral reason or principles is naive, Shklar argued, because an oppressive regime can use law to implement its oppression. There, government in accordance with the rule of law might involve the rigorous and impartial imposition of oppression but clearly that does not make such government legitimate.[3] The claim is

[1] Judith N. Shklar, "Political Theory and the Rule of Law", in Allan C. Hutchinson and Patrick Monahan (eds.), *The Rule of Law: Ideal or Ideology* (Toronto: Carswell, 1987), p. 1, all references below to this edition. (Shklar's essay is reprinted in Judith N. Shklar, *Political Thought and Political Thinkers* (Stanley Hoffman (ed.), Chicago: University of Chicago Press, 1998, p. 21.)

[2] Ibid., pp. 1–2.

[3] Ibid., pp. 13–14. Indeed one might suppose, though Shklar does not explicitly make this argument, that indiscriminately imposed injustice might be worse than arbitrarily imposed injustice.

parochial because it supposes that any particular legal order includes a judiciary empowered to test legislation in accordance with a foundational constitutional document such as the Bill of Rights.[4]

Shklar still seemed to regard the rule of law as potentially valuable, because it at least guards against arbitrary power. Even an oppressive regime which observes the rule of law has to show that its oppression has a legal warrant. And she argued that the idea that government has to justify its actions as being in accordance with the law has a properly valuable role to play in systems of representative democracy and of constitutional government. She concluded:

> "If one then begins with the fear of violence, the insecurity of arbitrary government and the discriminations of injustice one may work one's way up to finding a significant place for the Rule of Law, and for the boundaries it has historically set upon these the most enduring of our political troubles. It is as such both the oldest and the newest of the theoretical and practical concerns of political theory".[5]

Shklar's line of argument places her firmly within the positivist tradition of legal theory, one which is sceptical of any claim about the legitimacy of law and hence any claim about the inherent worth of making judges the guardians of the principles of the rule of law.[6] It is partly because the second claim seems always to follow hard on the heels of the first that the role of judges in the legal order has always been controversial.

The camps which divide on this question are, roughly speaking, democratic positivists and liberal anti-positivists. Democratic positivists, following the tradition established by Jeremy Bentham, argue that the legislature is the sole source of law and that its legitimacy derives from its accountability to the people. For judges to claim that the law is anything but the law enacted by the legislature is, therefore, for them to act undemocratically. In order for judges to fulfil their role in legal order of enforcing the will of the people, that will had to be expressed in legislation which made it as clear as possible what the content of that will is. Put differently, lack of ambiguity is what made judicial deference to the will of the legislature possible. Conversely, ambiguity or alleged ambiguity in the law gave judges the occasion for judicial legislation, and thus is best avoided. Positive law is, then, the law of the legislature which has the attributes which enable judicial deference to legislative will.

Bentham was a great opponent of the common law, even advocating its abolition. His opposition was driven by more than his sense that the common law was too messy to ever have the attributes of positivity. He was also concerned

[4] Judith N. Shklar, "Political Theory and the Rule of Law", n. 1 above, pp. 14–16.

[5] Ibid., p. 16.

[6] It is instructive to compare her essay with Joseph Raz, "The Rule of Law and Its Virtue", in Raz, *The Authority of Law: Essays on Law and Morality* (Oxford: Clarendon Press, 1983), p. 210. Although Raz's treatment of the rule of law is one from the perspective of a legal philosopher, while Shklar deals with the rule of law from the perspective of political theory, the substantive arguments are almost exactly the same.

about what he saw as the judicial device in a common law system of alleging that ambiguity existed in legislation in order to superimpose the judges' sense of right and wrong on the legislation. In other words, the common law provides a resource to judges which they can use to bootstrap themselves to the apex of legal order.

Anti-positivists, following a tradition most famously articulated by Sir William Blackstone, argue that the common law is not a mess but the legal repository of the moral values of the people. Judges, in enforcing common law values, are, on this view, giving effect to the will of the people. In using the common law as the value-laden background against which legislation is to be interpreted, judges are not setting themselves against the people's will because that background, no less than legislation, is the product of the people. Since proponents of this view identify common law values with the rights and liberties of the individual, we can therefore refer to them as liberal anti-positivists.

The division into democratic positivists and liberal anti-positivists is rough because the former set great store by the rights and liberties of the individual while the latter generally acknowledge that where a statute speaks clearly it legitimately overrides the common law. Nevertheless, the camps represent the poles on a continuum between which debate still moves today. And this is the case even in legal orders where fundamental values are constitutionalised. There, positivists will often argue for the existence of a determinate original content to the constitution against the liberal anti-positivist claim that the values have to be filled in by a process of judicial interpretation of the higher law, a process which is only framed by the text.

However, while the issue of whether the rule of law is necessarily the rule of moral principles—the issue of the legitimacy of law—is still central to debate about the rule of law, other issues have complicated that debate. This collection of essays by lawyers, philosophers, and political theorists illuminates and takes forward both the central debate about law and morality—the debate about the principles of the rule of law—and the debate about the issues which are currently complicating an already complex problem.

I have divided the essays into three main sections. The first contains studies of legal orders where the rule of law is under severe stress. These essays are united by their focus on the question of what one might conclude from such a study about the idea of the rule of law. The second section contains essays which engage the question of the value of the rule of law as a conceptual problem within philosophy of law, political philosophy, and legal theory. The third section contains essays which focus on the rule of law as a question about the limits of legal order. That is, they ask whether the rule of law has any application in areas where it might be thought to be only minimally effective—in transitions from dictatorship to democracy, in the control of the process of globalisation, and in the control of administrative agencies and executive discretion. What follows is a brief description of each section and of the contributions to it.

I. LAW UNDER STRESS

The anti-positivists who claim that the law is to some extent legitimate have always had to deal with the counter-example of the wicked legal system, the legal order which appears to have all the formal trappings of legality but which is the instrument of oppression. Contemporary debate abut this issue began when H.L.A. Hart and Lon L. Fuller clashed over Gustav Radbruch's claim that a prevailing positivism among German lawyers had helped to pave the way for National Socialism.[7]

Hart, one of this century's leading positivist thinkers, is generally thought to have prevailed over both Radbruch and Fuller in his argument that positivism, properly understood, offered a basis to lawyers and citizens for dealing candidly with injustice. However, the chapters in this volume by Robert Alexy and Julian Rivers go a long way to unsettling that thought.

Alexy deals with the problems faced by the German courts both when they evaluated the National Socialist past and when they dealt with the wall shooting cases after reunification, while Rivers focuses more on the wall shooting cases. Both conclude that these problems might be better dealt with by an anti-positivist theory of the rule of law. However, Alexy's version of anti-positivism depends on a more substantive account of the morality of law, while Rivers opts for a procedural account, more like Fuller's.

National Socialism is only one example of a wicked legal system, the legal order of apartheid is another standard reference point in this debate. Here Richard Abel's chapter, which summarises his book on law in the struggle against apartheid,[8] points out that the law, despite the efforts of a determined and powerful regime which used the law as an instrument of oppression, made a surprising amount of room available for opposition. Abel, however, does not commit himself to a position in the positivist/anti-positivist debate, which leaves open the question of what theoretical conclusion one can draw from this story.

I have in earlier work tried to argue that the story supports anti-positivism, since it shows that where there is law there has to be some repository of moral principles, and that judges who did the most to use the law to constrain oppression had to be anti-positivists.[9] My argument is effectively analysed and criticised in this volume by Anton Fagan. In his chapter, Fagan seeks to show that nothing in the sophisticated contemporary versions of positivism commits it to the view of the rule of law which I criticised as positivistic. He also argues

[7] H.L.A. Hart, "Positivism and the Separation of Law and Morals", reprinted in Hart, *Essays in Jurisprudence and Philosophy* (Oxford: Clarendon Press, 1983), p. 49 and Lon L. Fuller, "Positivism and Fidelity to Law: A Reply to Professor Hart" (1958) 71 *Harvard Law Review* 630.

[8] Richard L. Abel, *Politics by Other Means: Law in the Struggle Against Apartheid* (New York: Routledge, 1995).

[9] David Dyzenhaus, *Hard Cases in Wicked Legal Systems: South African Law in the Perspective of Legal Philosophy* (Oxford: Clarendon Press, 1991).

against the assumption in my work (and one which underpins Alexy's and Rivers's chapters here) that the validity of a legal theory depends upon the desirability of its practical effects.

Anthony Sebok's chapter deals with yet a third wicked legal system problem, taking as his foil Robert Cover's classic anti-positivist treatment of the responses of abolitionist northern judges to the plight of fugitive slaves.[10] Sebok analyses the central figure in Cover's account of judicial positivism, Judge Lemuel Shaw. While Sebok finds that Shaw was indeed a positivist, he was a very sophisticated positivist, that is, not a law-and-order, Hobbesian, obey-the-laws-commanded-by-the-sovereign positivist. Because his chapter offers a qualified defence of positivism and of Shaw, Sebok finds himself in something of a dilemma at the end, since he also wishes to highlight some of the negative effects of Shaw's positivism. In my view, the central predicament Sebok uncovers in this chapter can usefully be used to reflect on Fagan's conclusions.

Similarly, Alon Harel's chapter, while distinctively positivist in its theoretical commitments, seems to draw practical results from those commitments. His focus is the rule of law in Israel, in particular the way in which Israel's Supreme Court, led by a powerful Chief Justice, has sought to fashion a constitutional power of review from a fairly slim legal basis. While in the context of apartheid, critics of apartheid assumed that the rule of law was best served by judges willing to conceive of their review authority more expansively in order to constrain state oppression, Harel, discussing the Israeli context, argues that the rule of law is better served by judicial modesty.

II. RECONCEIVING THE RULE OF LAW

In a recent essay on the rule of law, the distinguished public law scholar Paul Craig draws a distinction between "formal" and "substantive" conceptions of the rule of law.[11] According to Craig, formal conceptions of the rule of law do not pass judgment on the content of the law; rather they "address the manner in which the law was promulgated (was it by a properly authorised person, in a properly authorised manner, etc.); the clarity of the ensuing norm (was it sufficiently clear to guide an individual's conduct so as to enable a person to plan his or her life, etc.); and the temporal dimension of the enacted norm (was it prospective or retrospective, etc.)".[12] A substantive conception, in contrast, for example, Ronald Dworkin's, derives its theory of the rule of law from an account of justice. "On this view the rule of law is nothing more or less than a synonym for a rights based theory of law and adjudication."[13]

[10] Robert Cover, *Justice Accused: Antislavery and the Judicial Process* (New Haven: Yale University Press, 1975).

[11] Paul Craig, "Formal and Substantive Conceptions of the Rule of Law: An Analytical Framework" [1997] *Public Law* 466.

[12] Ibid.

[13] Ibid., 477–9, at 479.

Craig draws attention to the connection between formal accounts and legal positivism and also suggests that there is no "middle way" between formal and substantive accounts, that is, a way which breaks down the process/substance distinction yet does not reduce the rule of law entirely to a substantive theory of justice.[14] His view seems to be that one should not abuse the idea of the rule of law by claiming that governments are not acting in accordance with the rule of law when what one means is that they are not acting justly. Rather, and here he places himself in the positivist tradition, if one has a substantive critique—for example, that judges are behaving undemocratically, one should make that critique explicit. However, it seems to me that there is a middle way and here I will sketch what I take it to be, before indicating how the three chapters included in this section might support the same thought.

Craig is, in my view, right to connect the formal idea of the rule of law to the positivist tradition. However, he is not right in his implication that somehow that idea stands free of a theory of justice. Indeed, the very claim that the rule of law is best understood formally—detached from a substantive theory of justice—is deployed in his hands in order to make a substantive claim about the best way to conduct political and legal debate.[15]

And within the positivist tradition, the formal idea of law was at one time seen as even more tightly connected to a substantive theory of justice. In the hands of the most powerful exponent of legal positivism, Jeremy Bentham, the rule of law should be seen as the rule of formal law because law so conceived best serves what I will refer to as the legal culture of reflection.

Such a legal culture works in the service of a radical democratic theory, in which statute law accurately reflects through the medium of the legislature the preferences of the majority. The judicial duty is to apply the law as it is and the citizen's duty to obey the law as it is. For Bentham, law properly so called is statute law, the reflection of the results of public rational reason, while the common law is the subjective opinions of judges masquerading as public reason. Judges are therefore to be disempowered as far as possible by depriving their judgments of legal force, except as between the parties.

Such a legal culture, while one of authority and obedience, is nevertheless far from authoritarian. Bentham is clear that the culture of reflection is legitimate only if political and legal institutions are highly responsive to legislative reform in the light of citizens' experience of the effects of the law. Bentham's famous slogan that the good citizen is the one who obeys punctually and censures freely makes sense only because he also requires political and legal channels for effective criticism of the law.

A different conception of the rule of law sees law as the embodiment of a culture of neutrality. This conception comes about through the liberal attempt to

[14] Paul Craig, "Formal and Substantive Conceptions of the Rule of Law: An Analytical Framework", n. 11 above, 477 and 484–6.

[15] I discuss the status of such claims in *Legality and Legitimacy: Carl Schmitt, Hans Kelsen and Hermann Heller in Weimar* (Oxford: Clarendon Press, 1997), ch. 3, which focuses on Hans Kelsen.

use law to preserve a realm of principles safe from democracy. It is a culture of neutrality because the main criterion for illegitimate state action is that the state has acted non-neutrally in infringing the individual's right to decide for himself how to live. In contrast to Bentham, such liberals hold that the principles which together make up the culture of neutrality, whether at common or constitutional law, represent the essence of law. These principles are supposed to express the true voice of public reason, since they demarcate the limits of the state's legitimate scope for interfering with individual liberty. Statutes are an inferior form of law, the transient expressions of majority preference as to government policy, legitimate only so long as they do not run up against the judges' understanding of the limits of public reason. Dworkin's theory of the rule of law is of course the best example of an attempt to articulate the culture of neutrality.

The middle way resides in the last of the ideas of legal culture—the culture of justification.[16] As I understand it, the idea of the rule of law as the rule of a culture of justification shares with Bentham the thought that the primary mode of making law is legislation, so that the role that judges have in legal order is one derived from a theory of legislation rather than from a Dworkin-like theory of adjudication. And it shares with the culture of neutrality the notion that judges should have a role in protecting fundamental principles of law. But those principles do not have the kind of judicially fixed content which liberals desire—they are not principles with a content against which statutes or executive decisions must not offend if they are to be valid. Rather, they are principles which make internal to the law the ideals of the common law of judicial review—the ideals of participation and accountability.

Bentham subscribed to both these ideals, but he did not make them into principles of law, preferring to provide after-the-fact political controls on the content of law. If you didn't like the law, you could try to get it changed through institutional channels which ensured legislative responsiveness to criticism. While every democrat would welcome such an attractive institution, it is an impoverished theory of democracy which does not countenance controls which operate in the very determination of what law is.

This point is underscored when one takes into account that the form of law has changed in ways unanticipated by Bentham, and by Hobbes—the founder of the positivist tradition. That is, the form has changed in just those ways that required the development of the body of law that in common law jurisdictions is now called administrative law—the principles that govern the activity of the officials who do not merely implement the law, but to whom the law delegates both policy-making and interpretative roles.

For a democrat, this change in the form of law might seem to require, at the least, that opportunities for participation be built into legal institutions. The

[16] I owe the name to the late South African administrative law scholar, Etienne Mureinik. I describe his work in "Law as Justification: Etienne Mureinik's Conception of Legal Culture", (1998) 14 *South African Journal on Human Rights* 11.

process of legislation, completed now only at the points when the administration develops enduring and legally binding policy, must be one which affords opportunities to those whose rights and interests will be affected and determined by that policy to participate in policy-making. But in any case, the idea that law in its nature has a determinate content must fail in the face of the radical indeterminacy of law in an era in which legislation is enacted in the service of broad social goals whose development, concretisation, and application are delegated to vast bureacracies.

I put this point somewhat tentatively because a democrat can take another tack, which is to try to reinstate Bentham's ideal. Here the left can join the right in an attack on the regulatory/administrative state though not on the welfare state. This left approach argues for a return to formal law, though to formal law that expresses very clearly policy goals and so does not require administrative officials to develop the policy content of those goals.[17]

It is important here to be aware of an ambiguity in the idea of formalism, which marks a real difference between, say, the right-wing formalism of F.A. Hayek and the left-wing formalism now being developed by William E. Scheuerman.[18] Hayek's formalism does stress clarity in the law, but the law is not supposed to set goals for individuals but to establish the framework conditions that make it possible for individuals to set their own goals. His formalism is an articulation of the role of law from the perspective of the culture of neutrality,[19] and not from the perspective of the culture of reflection. In contrast, Scheuerman is much closer to Bentham, arguing for law that is clear because clearly stated goals can be implemented without inviting a contest over how best to interpret the goal, a contest which will be won by the powerful whether it takes place in court rooms or in agency decision-making processes.

What both versions of formalism share is the view that the the principal value of the rule of law is certainty. And it is worth noting that Hayek's argument in his classic essay on the rule of law is that any attempt to go beyond his understanding of formalism will necessarily create uncertainty in the law, since the attempt by the state to plan for citizens always results in the creation of bureaucracies with vast discretionary powers over the content of the law. In other words, both versions of formalism fear discretion—the opposite of certainty—though one—Scheuerman's left-wing version—is optimistic about the elimination of discretion in a revival of the welfare state.

[17] The best example here is the argument that a minimum wage sufficient to live on should be paid to every adult regardless of her or his income. Such a wage could replace many different forms of welfare assistance and its universal nature means that it requires neither administration nor policing.

[18] See F.A. Hayek, "Planning and the Rule of Law", ch. 6 in his *The Road to Serfdom* (Chicago: University of Chicago Press, 1994), p. 80; William E. Scheuerman, "The Rule of Law and the Welfare State: Toward a New Synthesis" (1994) 22 *Politics & Society* 195. The last essay is discussed by Henry Richardson in his contribution to this volume.

[19] Of course, Hayek's account of neutrality is much more stringent than Dworkin's.

Where I think one should part company from both Hayek and Scheuerman is on the idea that discretion is to be feared, perhaps more accurately on their characterisation of the situation of judgment about the content of the law as a situation of discretion. Here I adopt Dworkin's critique of the positivist understanding of judicial decision as the result of an exercise of discretion without endorsing the principle/policy distinction he deployed in order to make that critique.[20] For Dworkin's great contribution to legal theory is, in my view, not so much his claim that there is a basis of liberal principles to the law revealed through judicial interpretation. Rather, his contribution lies in the way in which he has illuminated the justificatory character of the rule of law—law is not only about setting clear goals but also about argument as to what those goals should be.

In other words, Dworkin has shown that an important part of the rule of law is the idea—internal to the law—of justification. He also tries to argue for a set of ultimate principles which are immanent in the law, which must guide judicial interpretation. One can accept his account of the process without accepting those principles. The aim is then to uncouple the justificatory part of his theory from the substantive liberal "neutralist" part and then to try to conceptualise and design legal institutions in general, not just the judiciary, in order fully to realise the culture of justification.[21]

I hope that it is not reading too much into the chapters in this section to suggest that all three of them might also be seen as part of the law-as-justification project, one which seeks to do what we saw Craig declared impossible—staking out a middle way which breaks down the process/substance distinction yet does not reduce the rule of law entirely to a substantive theory of justice.

The first chapter in this section is by Neil MacCormick, one of the principal figures in the legal positivist tradition since Hart gave that tradition new direction in the late 1950s. Perhaps more than any other legal theorist writing today, MacCormick has been attentive to the potential to integrate insights from rival camps in legal theory into his own position. His essay for this volume illustrates this trait nicely, as he argues that one will not understand the rule of law in terms of an exclusively static conception, one which sees the rule of law merely as the rule of rules. Rather, the rule of law has also to be understood as including law as a locus of argumentation about what law is.

While MacCormick does not deal explicitly with the positivist/anti-positivist debate about the rule of law, in my view his chapter is either a significant amendment to positivism or part of an attempt to transcend that debate. For while Hart's positivism did of course account for argumentation, it seemed to demote

[20] For the distinction, see Ronald Dworkin, "The Model of Rules I" in *Taking Rights Seriously* (London: Duckworth, 1977), p. 14, at p. 22.

[21] It seems to me that this project is different from Jürgen Habermas's, outlined in *Between Facts and Norms: Contributions to a Discoourse Theory of Law and Democracy* (William Rehg (trans.), Cambridge, Mass.: MIT Press, 1996) because of the distinction between justification and application which he makes central to his legal theory. Habermas's account of the rule of law is briefly sketched in John P. McCormick's contribution to this volume.

the law which is arguable to a lesser status than determinate law by making the rule of law exclusively the rule of determinate law. However, for MacCormick, arguable law has the same status as determinate law, since he maintains that law is only defeasibly determinate or certain and that defeasability is as much a defence against arbitrariness as certainty.[22]

In her chapter, Christine Sypnowich argues explicitly for an understanding of the rule of law which will resolve the impasse between legal positivists and natural lawyers, one which will connect the proceduralism inherent in the idea of the rule of law with substantive standards of justice, but not so substantive as to amount to a liberal idea of neutrality. But this argument is not the exclusive focus of her essay, for, like Judith Shklar, her interest in the rule of law is driven more by the concerns of political than of legal philosophy. Sypnowich's concerns are to meet critiques from the left—critiques by democrats, feminists, and socialists who see the rule of law as a constraint on progressive social change.

One of the left critiques of the rule of law discussed by Sypnowich was advanced in a (co-authored) essay some years ago by Allan Hutchinson.[23] If Sypnowich's charge is accurate that Hutchinson at that time relied on a crude depiction of the rule of law, one imprisoned within the historical circumstances of its origin, it is clear from the next chapter in this volume that the target has moved. For here Hutchinson argues for a much more positive view of the role of courts in sustaining democracy than he had allowed in his earlier work. And this positive view is predicated on a sense which Hutchinson clearly shares with Sypnowich—that any firm process/substance distinction is unproductive if the rule of law is to be successfully reconceived.

III. THE LIMITS OF LEGAL ORDER

At a time when legal order could reasonably be conceived as made up of a separation of powers between the legislature, the judiciary, and the executive, one could agree that it was the task of the legislature to make law, of the executive to implement the law, and of judges to ensure that the executive stayed within the bounds of the law. Opinion divided on the question whether the legislature was the sole source of law or whether it was answerable to principles of a higher law, instantiated in the common law.

In the public law model of England and in those legal orders which follow the English model, the most influential understanding of the rule of law remains that put forward in 1885 by Albert Venn Dicey.[24] Dicey saw the main threat to

[22] Notice that on Robert Alexy's account in his chapter of the distinguishing features of positivism and anti-positivism, this would make MacCormick something of an anti-positivist. (And MacCormick does rely in his chapter on Alexy's theory of legal argumentation.)

[23] Allan Hutchinson and Patrick Monahan, "Democracy and the Rule of Law", in Hutchinson and Monahan, *The Rule of Law: Ideal or Ideology*, n. 1 above, p. 97.

[24] For an instructive account of Dicey, see P.P. Craig, *Public Law and Democracy in the United Kingdom and the United States of America* (Oxford: Clarendon Press, 1990), ch. 2. Judith Shklar

the rule of law in the development of administrative agencies. These, because they were often a curious hybrid of law-making and adjudicative functions, could not be neatly fitted into the separation of powers between legislature, courts, and an executive or administration supposedly confined to implementing determinate law.

Since the courts are supposed on this model to have an interpretative monopoly while parliament enjoys a legislative or law-making monopoly, judges who worked with the model had two options. Either they could adopt a hands-on approach and seek to confine the agencies to activity which looked more like law-application, or they could take a hands-off approach and simply declare that administrative agencies are not subject to the rule of law.

Neither option is satisfactory since the hands-on approach applies a model of the rule of law to contexts where it is unsuited while the hands-off approach leaves those subject to official discretion without the protection of the rule of law. Moreover, when judges attempted to implement the hands-on option, often in order to hold back the growth of a welfare state to which they were politically opposed, legislatures exercised their legislative monopoly and told courts through the device of the privative or ouster clause that they had no jurisdiction over agencies. In England, judges simply read such clauses out of statutes while in other jurisdictions, for example, Canada, judges eventually tried to work out a new, more positive relationship between administrative agencies and courts. In the context of common law courts' struggles to understand the role of administrative agencies we see graphically the problem of the limits of legal order or of the reach of the rule of law.[25] But this problem is now being raised even more graphically in other contexts.

One is the context of transitions to democracy from dictatorial regimes, where the rule of law was at best a facade, to democracy. Here the question is what role the law and the rule of law can play in ensuring a successful transition from a situation where law was used as an instrument of oppression. In his chapter in this volume, Richard Abel suggests that the fact that law could be used to resist legalised oppression in South Africa means that law and the rule of law provide a useful resource for managing the transition there.[26] But one might not suppose that the rule of law will always be so helpful after reading Andras Sajo's essay on the rule of law in post-communist Hungary. Written

said that Dicey's restatement of the rule of law "was the most influential" since the eighteenth century, though in her view this influence was regrettable as Dicey's restatement amounted to an "unfortunate outburst of Anglo-Saxon parochialism"; Shklar, "Political Theory and the Rule of Law", n. 1 above, p. 5.

[25] For my vesion of the Canadian story, see "The Politics of Deference: Judicial Review and Democracy" in Michael Taggart (ed.), *The Province of Administrative Law* (Oxford: Hart Publishing, 1997), p. 279. As the title of the book suggests, the common theme of the essays is the question of the reach of the rule of law.

[26] See more generally, A. James McAdams (ed.), *Transitional Justice and the Rule of Law in New Democracies* (Notre Dame: University of Notre Dame Press, 1997) and, in regard to South Africa, my *Judging the Judges, Judging Ourselves: Truth, Reconciliation and the Apartheid Legal Order* (Oxford: Hart Publishing, 1998).

from an intensely personal point of view, Sajo describes a realm of rampant freedom—a spiral of disorder—which follows the breakdown in communist order.

A third context which dramatically exposes problems to do with the reach of the rule of law is globalisation. As multinational corporations increasingly take control of national economies, and as economic activity becomes increasingly controlled through the lightning-fast Internet, it becomes very difficult to see how actors who have a tremendous impact on our lives can be thought of as subject to the rule of law.

In his essay, William E. Scheuerman exposes the different aspects of this problem, all of which go to show that the link often supposed to exist between the rule of law and capitalism has to be rethought. Scheuerman does not, however, give up the hope that globalisation could be subjected to the rule of law. And in his essay, John P. McCormick explores the possibility that one supranational organisation, the European Union has, through the European Court of Justice, begun to establish a model for a supranational rule of law which will not necessarily subvert social democracy.

The last three chapters focus on the context with which I began the present discussion—the control of administrative discretion. Kenneth Winston's chapter deals with the conceptions of the rule of law at stake in the exercise of discretion by a political appointee to high executive office in the United States of America—the Solicitor General. In his close analysis of the office when it was held by the distinguished legal scholar Charles Fried, Winston continues his project of elaborating Lon L. Fuller's contribution to legal theory.

In his chapter, Henry S. Richardson sets out the philosophical case for supposing that, properly understood, the exercise of administrative discretion is compatible with both the rule of law and democracy. In particular, he argues that that there is no need to be overly concerned by the phenomenon of representation of special interests in decision-making. Part of his argument criticises recent work by Jody Freeman, and Freeman's chapter in this volume responds to that critique. She argues that Richardson has not identified the real problem of democracy at stake in the administrative context—the problem of how to conceive of the public realm given that both public and private actors interact in it—and she elaborates a model for dealing with that problem.

Part I

Law Under Stress

2

A Defence of Radbruch's Formula[1]

ROBERT ALEXY

The problem of dealing with a past devoid of the rule of law has twice confronted the courts in Germany in a century which is now drawing to a close: in 1945 after National Socialism was crushed and in 1989 after the collapse of the German Democratic Republic. In both cases the following question had to be answered. Should one regard as continuing to be legally valid something which offended against fundamental principles of justice and the rule of law when it was legally valid in terms of the positive law of the legal system which had perished. To use a handy though imprecise formulation, can something be illegal today which in the past was legal? After 1945 German courts answered "yes" to this question,[2] and the Federal Court of Justice has followed this tradition after 1989 especially in its decisions in regard to the so-called wall shootings.[3] The Federal Constitutional Court forged ahead in this course in cases of National Socialist injustice[4] and affirmed it in cases concerning the injustice committed by the German Democratic Republic.[5] Radbruch's formula formed the jurisprudential core of the judges' reasoning. In what follows, the first task will be to present this formula. Then its practical significance will be illustrated through two examples. Finally, we will ask whether the formula can stand up to jurisprudential critique.

I. THE FORMULA

Gustav Radbruch presented his famous formula under the immediate impression of twelve years of National Socialism. It reads:

> "The conflict between justice and legal certainty should be resolved in that the positive law, established by enactment and by power, has primacy even when its content

[1] Translated by David Dyzenhaus. I thank Professor Alexy for the great care he took in suggesting improvements to my first draft though the responsibility for all errors is mine.

[2] See for example OGHSt 2, 231 (232 ff.); 2, 269 (272 ff.); BGHSt 2, 173 (177); 2, 234 (237 ff.); 3, 357 (362 ff.).

[3] BGHSt 39, 1; 39, 168; 39, 199; 39, 353; 40, 48; 40, 113; 40, 218; 40, 241; 41, 10; 41, 101; 41, 149; 42, 65; 42, 356.

[4] BVerfGE 3, 58 (119); 3, 225 (232 ff.); 6, 132 (198); 6, 389 (414 ff.); 23, 98 (106); 54, 53 (67 ff.).

[5] BVerfGE 95, 96 (130 ff.).

> is unjust and improper. It is only when the contradiction between positive law and justice reaches an intolerable level that the law is supposed to give way as a 'false law' [*unrichtiges Recht*] to justice. It is impossible to draw a sharper line between the cases of legalized injustice and laws which remain valid despite their false content. But another boundary can be drawn with the utmost precision. Where justice is not even aimed at, where equality—the core of justice—is deliberately disavowed in the enactment of a positive law, then the law is not simply 'false law', it has no claim at all to legal status".[6]

It is easy to see that this formula is composed of two parts. In the first part, the claim is that positive law loses its legal validity when its contradiction with justice reaches an "intolerable level". We can call this the "intolerability formula". In the second part, positive laws are denied legal status when in their enactment equality, which Radbruch says is the core of justice, is "deliberately disavowed". We can call this the "disavowal formula".[7] The intolerability formula has an objective character. It is attuned to the level of injustice.[8] In contrast, there is something subjective about the disavowal formula: the purpose or intentions of the legislators. One can think of cases in which both formulae lead to different results. One can easily imagine a legislator who in fact strives for equality as the core of justice, but nevertheless brings about something which is intolerably unjust, just as one who is bent on bringing about injustice might fail to cross the threshold of intolerable injustice. But in general it is true that result and purpose should coincide when in issue is intolerable injustice. In this repect one can speak of an "overlapping" of both formulae.[9] Judicial reasoning has first and foremost applied the intolerability formula. In favour of this is that an "intention to warp justice" is very difficult to prove in doubtful cases.[10] In this chapter, the intolerability formula is the focus.

Remarkable about Radbruch's formula is that it does not require any complete coincidence between law and morality. It allows enacted and effective law—Radbruch speaks of the "law established by enactment and by power"—to be valid even when it is unjust and it does not even require that the law as a whole orient itself to morality. It is much more the case that it builds into law an outermost limit. In general, law is that which is appropriately enacted and

[6] G. Radbruch, "Gesetzliches Unrecht und übergesetzliches Recht", in G. Radbruch, *Gesamtausgabe*, in A. Kaufmann (ed.), (Heidelberg: C.F. Müller, 1990), vol. 3, p. 89. Radbruch's famous essay was first published in 1946 in the first volume of the *Süddeutschen Juristen-Zeitung*, 105–8.

[7] See B. Schumacher, *Rezeption und Kritik der Radbruchschen Formel* (Diss: Göttingen, 1985), p. 24 ff.; A. Kaufmann, "Die Radbruchsche Formel vom gesetzlichen Unrecht und vom übergesetzlichen Recht in der Diskussion um das im Namen der DDR begangene Unrecht" (1995) 48 *Neue Juristische Wochenschrift* 81 at 82.

[8] More precisely, two aspects should be distinguished within the framework of the intolerability formula. The first concerns the weighing, the second the threshold.

[9] S.L. Paulson, "Radbruch on Unjust Laws: Competing Earlier and Later Views?" (1995) 15 *Oxford Journal of Legal Studies* 489 at 491.

[10] R. Dreier, "Gesetzliches Unrecht im SED-Staat? Am Beispiel des DDR-Grenzgesetzes" in F. Haft, W. Hassemer, U. Neumann, W. Schild, U. Schroth (eds.), *Strafgerechtigkeit. Festschrift für Arthur Kaufmann* (Heidelberg: C.F. Müller, 1993), p. 58.

socially effective; only when the threshold of extreme injustice is crossed do appropriately enacted and socially effective norms lose their legal character or their legal validity. Hence, one can express Radbruch's formula concisely:

> appropriately enacted and socially effective norms lose their legal character or their legal validity when they are extremely unjust.

Even shorter:

> Extreme injustice is no law.[11]

Whoever supports this thesis has ceased to be a legal positivist. When a positivist wants to establish what law is, he inquires only into what is appropriately enacted and socially effective. Though these are ideas which can be very differently interpreted and evaluated, as the many forms of legal positivism show, nothing more will be said on this issue here. Of interest in this chapter is only that for the positivist nothing about legal character or validity turns upon the content of the norm. The great legal positivist Hans Kelsen expressed this idea in a much cited formulation: "Hence any content whatsoever can be legal".[12] This is the positivist thesis of the separation of law and morality, in short, the positivist separation thesis. Even the anti-positivist takes into account appropriate enactment and social effectiveness if he wishes to be regarded as in his right mind. Radbruch's formula is clear evidence of this. But for the anti-positivist who adopts the formula there is nevertheless a limit, that of extreme injustice. In this way substantive correctness is imported as a limiting criterion into the concept of law. The concept of law is not filled out by morality but it is limited by morality. This is clearly only a partial connection of law and morality but it is a connection. Whoever advocates Radbruch's formula therefore supports the anti-positivist connection thesis.[13]

The conflict about Radbruch's formula is a philosophical conflict because it is a conflict about the concept of law. It speaks volumes about the character of legal philosophy that this conflict over its foundational concept—the concept of law—has at the same time direct practical consequences. We will take such consequences into account before we ask whether the better argument speaks for or against anti-positivism in the form of Radbruch's formula. And this can be done through two examples.

[11] Radbruch comes close to this formulation when he says that "horribly" unjust laws can be denied validity; see G. Radbruch, *Vorschule der Rechtsphilosophie*, in G. Radbruch, *Gesamtausgabe* (A. Kaufmann (ed.), Heidelberg: C.F. Müller, 1990), vol. 3, p. 154. The *Vorschule der Rechtsphilosophie* was published first in 1948 as lecture notes which were revised and approved by Radbruch.

[12] H. Kelsen, *Reine Rechtslehre*, 2nd edn. (Vienna: Deuticke, 1960), p. 201.

[13] R. Alexy, *Begriff und Geltung des Rechts*, 2nd edn. (Freiburg and Munich: Alber, 1994), p. 15 ff., p. 52 ff.

II. PRACTICAL SIGNIFICANCE

The first example is a 1968 decision of the Federal Constitutional Court concerning Decree 11 in regard to the Reich's Citizenship Law of 25 November 1941.[14] Section 2 of this decree reads:

> "A Jew loses German nationality
> (a) with the coming into force of this decree when he has his usual residence abroad at the time of the coming into force of this decree,
> (b) when he at a later date takes up his usual residence abroad at the time when he changes his usual residence to abroad".

The occasion of the Federal Constitutional Court's decision was whether a Jewish lawyer, who had emigrated to Amsterdam shortly before the Second World War, had lost his German nationality in accordance with this rule. The outcome of a matter concerning an inheritance turned on this point. The lawyer had been deported from Amsterdam in 1942. There was no news about his fate beyond that, so it had to be accepted that he had lost his life in a concentration camp.

The Federal Constitutional Court decided that the lawyer had not lost his German nationality because Decree 11 in regard to the Reich's Citizenship Law was from the outset void. The core of its reasoning reads:

> "Hence the Federal Constitutional Court has affirmed the possibility of depriving National Socialist 'legal' decrees of their legal validity because they so evidently contradict fundamental principles of justice that the judge who applied them or recognised their legal consequences would pronounce injustice instead of law (BVerfGE 3, 58 (119); 6, 132 (198)).
>
> Decree 11 offends these fundamental principles. In it the contradiction with justice has reached so intolerable a level that it must be regarded as void from the outset (see BGH, RzW 1962, 563; BGHZ 9, 34 (44); 10, 340 (342); 16, 350 (354); 26, 91 (93))".[15]

This is a classical anti-positivist argument. An appropriately enacted norm, one which was socially effective for the duration of its validity, is denied validity or—on this point the decision is not unequivocal—its character as law, because it offends suprapositive law. While Radbruch was not in fact mentioned by name, one finds his name nevertheless in earlier decisions of the Federal Constitutional Court on which the Court in this decision expressly relied.[16] In any case more significant is that Radbruch's formulation of the "intolerable level" of the "contradiction" with "injustice" is applied. The decision on nationality is thus a paradigmatic case of the application of Radbruch's formula.

Expatriates often had no desire to get their old citizenship back. But generally in the case of property things were different. This was the issue in a decision of

[14] RGBl. I p. 722.

[15] BVerfGE 23, 98 (106).

[16] In the decision BVerfGE 3, 225 (232), which is mentioned in BVerfGE 22, 98 (106) immediately prior to the text quoted above, Radbruch's formula is cited in its entirety, word for word.

the Great Panel of the Federal Court of Justice for Civil Matters, which should rank with the decision on nationality. Once again the outcome of the proceedings turned on the validity of Decree 11 in regard to the Reich's Citizenship Law, this time on section 3, paragraph 1, provision 1, which reads:

> "The property of Jews who have lost their German nationality on the basis of this Decree becomes the property of the Reich with the loss of nationality".

A Jewish woman who emigrated to Switzerland in 1939 had left securities in a deposit in a German bank. During the entire period of National Socialist rule and also thereafter this deposit remained entered in the books of the bank in the name of the emigrant. After the end of the war she again took up domicile in the Federal Republic of Germany. Presently she demanded that the securities in the deposit be restored to her. The question was whether she had lost her property on the basis of the immediate expropriation in terms of section 3, paragraph 1, provision 1 of Decree 11. The Federal Court of Justice answered "no" to this question and therefore confirmed her demand for restitution. The details of its reasoning are complex but the core reads as follows:

> "§3 of Decree 11 under the Reich's Citizenship Law is to be regarded as from the outset void because of its iniquitous content which contradicts the foundational requirements of every order based on the rule of law".[17]

Following this anti-positivist solution the emigrant could demand her property back simply because she had never lost it. From the standpoint of legal positivism some retroactive or correcting regulation was required if she were to have any title to claim restitution. Whether she could demand the property back would hinge then on the discretion of the legislature. The decision for or against legal positivism therefore can have immense practical significance for the victim of a tyrannical regime.

The second example of the practical significance of Radbruch's formula comes from the judicial decisions in regard to the deaths of fugitives on the border formerly internal to Germany. The Federal Court of Justice confirmed the guilt of simple border soldiers in its first judgment on wall shootings in November 1992,[18] a good two years after reunification. Two years later in 1994 it decided that higher and the highest German Democratic Republic officials, among them the last Minister of Defence of the German Democratic Republic, Army-General Keßler, were criminally responsible for the killings on the border. It found them guilty of being the indirect cause of manslaughter.[19] Again two years later, in October 1996, the Federal Constitutional Court declared this line of adjudication to be in accordance with the Constitution.[20] Here only the leading decision is examined—the first judgment on wall shootings by the Federal Court of Justice.

[17] BGHZ 16, 350 (354).
[18] BGHSt 39, 1.
[19] BGHSt 40, 218.
[20] BVerfGE 95, 96.

This judgment concerned a twenty-year-old fugitive, who on 1 December 1984 at about 3.15 a.m. attempted to get over the border structure with a four-metre-long ladder. Two soldiers of the border patrol of the German Democratic Republic, one about twenty the other about twenty-three years old, caught sight of the fugitive about 100 metres away, as he prepared to cross the 29-metre-wide border strip. In the middle of the strip stood a 2.5-metre-tall alarm fence and at its end stood a 3.5-metre-high border wall. Neither calls nor warning shots could stop the fugitive. As he leant his ladder against the border wall and quickly ascended, it became clear to the two soldiers that only directed fire stood any chance of preventing his flight. They shot several bursts of fire at the fugitive. Though they aimed at his legs, they knew that there was the possibility that he would be killed especially because of their sustained fire. But even at this price they were determined to prevent his flight. The fugitive was hit a few seconds after they opened fire, at the moment his hand reached the top of the wall. He died within hours.

In 1992 the Berlin Provincial Court found both soldiers jointly guilty of manslaughter and sentenced the younger one to imprisonment in a young offender's facility for a year and six months and the older to a prison term of one year and nine months.[21] The execution of both punishments was deferred pending probation. In its first wall shooting decision, the Federal Court of Justice rejected the appeals against this judgment and confirmed the convictions though not the reasoning behind them.

In accordance with the rules of the treaty on the restoration of German unity the general principle was valid for both soldiers that their deed was punishable only if it was punishable in terms of the valid law governing at the time and in the place it was done. The crucial question was thus whether they had a permission or justification in terms of the law then in force in the German Democratic Republic. In issue as their ground of justification was section 27 of the 1982 Border Law of the German Democratic Republic.[22] In the present case, section 27, paragraph 2, provision 1 was significant:

> "The use of firearms is justified to prevent the directly imminent carrying out or the continuation of a criminal act which, in the circumstances, appears to be a felony".

The fugitive's crossing of the border was directly imminent and the soldiers fired to prevent him from that. According to the interpretation of the criminal law—both the dominant theory and the practice—of the German Democratic Republic it was a felony to attempt to break through the border as the fugitive had done.[23] Hence all the preconditions of section 27, paragraph 2, provision 1 were in evidence. Even the remaining preconditions of section 27 were fulfilled.

[21] LG Berlin, NStZ 1992, 492 (493). Army-General Keßler has incurred the most severe punishment so far, as he was sentenced to seven years and six months; see BVerfGE 95, 96 (97).

[22] DDR-GBl. I p. 197.

[23] See R. Alexy, *Mauerschützen. Zum Verhältnis von Recht, Moral und Strafbarkeit* (Hamburg: Vandenhoeck & Ruprecht, 1993), p. 11 ff.

Fire was only opened after milder measures did not work (section 27, paragraph 1, provision 2). In this case the fugitive could only be stopped by fire. He was called back and a warning shot was fired (section 27, paragraph 3). Finally, section 27, paragraph 5 had to be observed:

> "When firearms are used the life of the person is if possible to be spared".

Even this norm was respected since it does not require that one may not in any way threaten life. It only says that "if possible" life is to be spared. The flight could not have been prevented at all without firing at the fugitive and, given that he was seconds away from succeeding, single shots would not have been as sure. When the prevention of the flight is understood as a justificatory ground in the sense of section 27, paragraph 2, it follows that section 27, paragraph 5 was also not violated.

The attempt to convict the border soldiers was undertaken by interpreting section 27 of the Border Law of the former German Democratic Republic in the light shed in the present by the principles of the rule of law. The judgment of the Berlin Provincial Court, which convicted both soldiers in the first instance, is an example of this. It held that the soldiers should have complied with the fundamental principle of proportionality, which meant that the soldiers should not have opened continuous fire. In addition, the aim of preventing a criminal act which did not endanger the life of another could never justify the killing of a person, since life is the most prized legal value.[24]

One should welcome the fact that the Federal Court of Justice did not adopt this reasoning, at least in the first part of its judgment which is the part of interest here. Whoever interprets the former law of the German Democratic Republic in the light shed in the present by principles of the rule of law is pursuing a covert kind of retroactivity which is worse than an open one. The question whether the punishment today of both soldiers offends the proposition *Nullum crimen sine lege* or *Nulla poena sine lege* would be evaded. In this regard, the Berlin Provincial Court got the positive law wrong. For not only the wording of the positive law makes up the positive law in force at the time; there is also the interpretative practice of the time. If one applies this standard, then the deed of both soldiers, as the Federal Court of Justice effectively and in full detail showed,[25] was justified by section 27, paragraph 2, provision 1 of the Border Law of the German Democratic Republic. The deed was thus legal in terms of the positive law valid at that time. Since a retroactive law which declared the deed as punishable today did not exist, both soldiers could be punished only if the justificatory ground in section 27, paragraph 2, provision 1 did not apply. The Federal Court of Justice brought Radbruch's formula into play on exactly this point:

> "It is much more the case that a justificatory ground taken from the time of the deed can only be disregarded because of its offence to a higher order of law when in it is

[24] LG Berlin, NStZ 1992, 492 (494).
[25] BGHSt 39, 1 (10 ff.).

> manifested a patently gross offence to the fundamental tenets of justice and humanity; the offence must be so weighty that it violates the legal convictions of all nations in regard to people's worth and dignity (BGHSt 2, 234, 239). The contradiction between positive law and justice must be so intolerable that the law has to give way to justice as a false law (Radbruch, SJZ 1946, 105, 107)".[26]

The last sentence is an almost word for word repetition of Radbruch's intolerability formula. The Court then explained that the scope of application of Radbruch's formula was not limited to National Socialist injustice:

> "In these formulations (see also BVerfGE 3, 225, 232; 6, 132 f., 198 f.) the attempt was made to define the worst violations of the law after the end of the National Socialist regime of violence. The transfer of these perspectives to the present case is not easy, because the killing of people on the internal German border cannot be equated with National Socialist mass murder. All the same, the insight achieved at that time remains valid that in judging deeds done at the command of the state one has to take into account whether the state has crossed the outer limits which are ordained to it by general convictions in any country".[27]

Thus everything turns on the question whether the deaths on the internal German border amounted to an extreme injustice in Radbruch's sense. This is very controversial.[28] The Federal Court of Justice answered in the affirmative with a detailed justification in the guarantees of the rights to life (Article 6) and mobility (Article 12) of the International Covenant on Civil and Political Rights of 19 December 1966, which, as it said, were drawn on as "guiding principles". This reasoning will not be reiterated here, since the issue is the presentation of the way in which Radbruch's formula works in practice. And this is shown with great clarity in the sentence with which the Federal Court of Justice removed from both border soldiers the justification in terms of the positive law of the former German Democratic Republic:

> "The justification stipulated by the law of the German Democratic Republic, described in § 27 of the Border Law, had for this reason from the outset no validity in the interpretation which is defined by the actual relations on the border".[29]

III. THE ASPIRATION AND THE LIMITS OF LAW

Radbruch's formula excludes certain contents from entering into the content of law, namely extreme injustice. In this way it restores a necessary connection between law and morality, that is, between the law as it is and the law as it ought to be. Appropriately enacted and socially effective law does not, to be sure, have to be just or right in order to be law, but it must not cross the threshold of

[26] BGHSt 39, 1 (15 ff.).
[27] BGHSt 39, 1 (16).
[28] See R. Alexy, *Mauerschützen*, n. 23 above, p. 23 ff.
[29] BGHSt 39, 1 (22).

extreme injustice. If this occurs, its legal character or validity as law is lost. This is a denial of the positivist thesis that there is a complete separation of law and morality and a profession of the anti-positivist connection thesis.

1. A conceptual framework

The conflict over legal positivism seems to be a conflict with no end, and that means it is a philosophical debate. In such disputes which are at once endless, acute and stubborn, one can surmise that all the participants are right in one or other aspect or in regard to one or other assumption. Our next task will be to cast a glance over these aspects or assumptions and here four distinctions are useful.[30]

(a) Norm and procedure

The first distinction is between the legal system as a system of norms and the legal system as a system of procedures. As a system of *procedures* the legal system is a system of interactions dependent on rules and guided by rules by means of which norms are enacted, grounded, interpreted, applied and executed. As a system of *norms* the legal system is a system of results or products of the procedures provided for producing norms. This distinction approximates Fuller's between the law as "activity"[31] in the sense of a "purposive effort that goes into the making of law and the law that in fact emerges from that effort",[32] hence, the law as "product"[33] or "results".[34] It is obvious that the understanding of law as a system of procedures or activities is more suitable to an anti-positivist position than the exclusive focus on norms as the results of such procedures.

(b) Observer and participant

The second distinction is between the observer and participant perspectives. This dichotomy maps onto Hart's distinction between an "external" and an "internal" point of view.[35] Hart's distinction is clearly in need of interpretation.[36] Here

[30] For the sake of simplification, I will here avoid the distinction employed in earlier work between concepts of law which include the concept of validity and those which do not; see R. Alexy, "On Necessary Relations Between Law and Morality" (1989) 2 *Ratio Juris* 167 at 170. The list presented here could be supplemented with more distinctions than the one just mentioned. For example, it is very fruitful for some purposes to take up the distinction between single norms and legal systems as a whole; see R. Alexy, *Begriff und Geltung*, n. 13 above, p. 57 ff., 108 ff.

[31] L.L. Fuller, *The Morality of Law*, 2nd edn. (New Haven and London: Yale University Press, 1969), pp. 106, 119.

[32] Ibid., p. 193.

[33] Ibid., p. 106.

[34] Ibid., p. 119.

[35] H.L.A. Hart, *The Concept of Law*, 2nd edn. (Oxford: Clarendon Press, 1994), p. 89.

[36] See N. MacCormick, *Legal Reasoning and Legal Theory* (Oxford: Clarendon Press, 1978), p. 275 ff.

we will interpret it with the help of the concepts of argumentation and of correctness: the *participant perspective* engages one who within a legal system takes part in argumentation about what it requires, forbids and permits and in addition about what it enables. The judge stands at the centre of the participant perspective. When other participants, including legal academics, lawyers, and citizens who concern themselves with the legal system, bring forth arguments for or against the particular meaning of laws, then they refer ultimately to what a judge would decide when he wanted to make a correct decision. The *observer perspective* engages one who asks not what the correct decision is in a particular legal system, but what the actual decision in a particular system will be. Again it is easy to recognise that the observer perspective is more suitable for the positivist and the participant perspective for the anti-positivist.

(c) Classification and qualification

The third distinction concerns the two different kinds of connection between law and morality. The first kind will be defined as "classificatory", the second as "qualificatory". One has to do with a *classificatory* connection when one maintains that norms or systems of norms which fail to meet a particular moral criterion fail to be legal norms or legal systems. Radbruch's formula creates such a connection since it excludes legal norms containing extreme injustice from the class of legal norms (or of valid legal norms). One has to do with a merely *qualificatory* connection when one maintains that norms or systems of norms which fail to meet a particular moral criterion could indeed be legal norms or legal systems, but are legally defective legal norms or legally defective legal systems. It is crucial that the defect asserted is a legal one and not merely moral.

The concept of a qualificatory connection is tightly bound up with the claim to correctness, since if the law necessarily raises a claim to correctness, then there necessarily exists a qualificatory relationship between law and morality.[37] Fuller's "internal" or "inner morality of law" as a "morality of aspiration"[38] resembles in a crucial respect the thesis of the claim to correctness. The incomplete fulfilment of the eight "principles of legality" which, acccording to Fuller, define the "inner morality of law", do not lead in general to a loss of legal character or legal validity.[39] It has therefore no classificatory meaning but rather the result is a qualification of the law or legal system as "bad".[40] Thus Fuller's theory is a classic example of theory which essentially depends on qualificatory connections.

[37] See R. Alexy, "Law and Correctness" in M. Freeman (ed.), *Legal Theory at the End of the Millennium* (Oxford: Oxford University Press, 1998), p. 214 ff.

[38] L. Fuller, *The Morality of Law*, n. 31 above, p. 43.

[39] Ibid., pp. 39, 41 ff.

[40] Ibid., p. 39. In contrast, one has to do with a classificatory connection when Fuller says that "a total failure" in the fulfilment of any one of his eight principles of legality "does not simply result in a bad system of law; it results in something that is not properly called a legal system at all"; ibid.

The qualificatory connection does not imply any classificatory one.[41] It is however easier to justify the latter when the former exists than when it does not. The justification of Radbruch's formula will thus begin with the justification of the qualificatory connection.

(d) Analytical and normative arguments

The fourth distinction is that between analytical and normative arguments for and against legal positivism. An *analytical* argument is presented when one shows it to be the case that the inclusion of moral elements in the concept of law is conceptually or linguistically necessary, impossible or merely possible. In contrast, the separation or connection thesis is supported by a normative argument when it is proposed that the inclusion or exclusion of moral elements is necessary to fulfil certain norms, such as the prohibition on retroactivity, or to realise certain values, such as human rights.[42]

As we have already seen, Radbruch's formula has to do with a classificatory connection. That the issue here cannot be decided on analytical grounds alone is demonstrated by the fact that neither of the following two sentences contains a contradiction:[43]

(1) The Norm N is appropriately enacted and socially effective and therefore law even if it contains extreme injustice.
(2) The Norm N is appropriately enacted and socially effective but not law because it contains extreme injustice.

Because of the vagueness and ambiguity of the expression "law" (*Recht*), a decision on the correctness of Radbruch's formula is ultimately possible only on the basis of normative arguments. These lead to completely different results depending on whether one adopts the observer or the participant perspective.

[41] See N. MacCormick, "Natural Law and the Separation of Law and Morals", in R.P. George (ed.), *Natural Law Theory* (Oxford: Clarendon Press, 1992), p. 112 ff., p. 130.

[42] One could suggest that whoever justifies the positivistic thesis of the separation of law and morality with moral and hence normative arguments ceases to be a positivist. On this interpretation, it would be the case that any use of a moral argument in the framework of a theory of law turns the theory into an anti-positivist one. One reason that speaks against such an extremely strict definition of legal positivism is that hardly any positivist would survive. A much more weighty consideration is that the crucial difference is flattened between authors who claim that a norm loses its legal character or validity when it violates a moral criterion and authors who claim that nothing about legal character or validity turns on any moral criterion. Both supporters and opponents of Radbruch's formula could then be equally characterised as anti-positivists when they adduced for their position any non-positivist, normative and in this sense moral argument, for example, that of legal certainty. This mode of conceptual argument would be confusing.

[43] Hart suggests that "the positivist might point to a weight of English usage" that sentences like (1) contain no contradiction; see H.L.A. Hart, *The Concept of Law*, n. 35 above, p. 209. The argument is then to be extended to sentences like (2). Hart's argument is then compelling: "Plainly we cannot grapple adequately with this issue if we see it as one concerning the proprieties of linguistic usage".

2. The observer

To answer the question whether Radbruch's formula is acceptable from the observer standpoint we will return to Decree 11 in regard to the Citizenship Law of the Reich of 25 November 1941, which deprived the Jew who had emigrated of citizenship and property. Imagine a contemporary observer of the National Socialist legal system, a foreign jurist who was composing a yearly report on the National Socialist legal system for a law journal in his homeland. How would he at the end of 1941 describe the case presented above of the emigrant whose securities section 3, paragraph 1, provision 1 of that Decree declared to be the property of the German Reich? It would be the case that anyone in his homeland would understand without any further explanation the proposition:

(1) A lost ownership of the securities in accordance with German law.

But this is not the case with the following proposition:

(2) A did not lose ownership of the securities in accordance with German law.

When no further information is given with this second proposition, he is either giving false information or confusing. The reason is that one can use the expression "law" in a way which serves only the value neutral identification of appropriately enacted and socially effective norms together with their consequences. Only this use is appropriate for the observer perspective. It serves clarity and truth of speech. A lawyer who had to advise a Jew at the end of 1941 and neglected Decree 11 would be in gross dereliction of duty. An appeal to Radbruch's formula would not help him in any way. Naturally, he could conclude his opinion with the following remark:

(3) She has lost her property in accordance with regulations which are now valid in Germany, but which amount to extreme injustice and are therefore not law. After the collapse of National Socialism we will ensure that the loss of property is declared to be invalid.

With this, the position of the mere observer is relinquished and one takes up in anticipation the position of a participant in a discourse about how to classify legally the expropriation after the collapse of the dictatorship. With this change of perspective, the expression "law" takes on a different meaning.

3. The correctness argument

Properly understood the real question in the debate about Radbruch's formula is whether it is acceptable from the standpoint of a participant in a legal system. Here one has to distinguish between participants in legal procedures in an iniquitous state and participants in procedures which begin to come to terms with

the former injustice after the collapse of the system. The question of whether the law necessarily raises a claim to correctness plays a decisive role in an explication of what it means to be a participant in a legal system. The thesis that the law necessarily raises such a claim can be called the "correctness argument". The correctness argument makes up the basis of the justification of Radbruch's formula.

The correctness argument maintains as valid that individual legal norms, individual legal decisions, and also whole legal systems necessarily raise a claim to correctness.[44] This can be demonstrated by examples in which there is an explicit negation of the claim to correctness. Here only one is dealt with.[45] It concerns the first provision of a new constitution for state X, in which a minority suppresses the majority. The minority would like to enjoy the advantages of the suppression of the majority while being honest about it. Their constitutional assembly thus decides on the following as the first provision of the constitution:

X is a sovereign, federal and unjust Republic.

Something is flawed in this constitutional provision, but the question is in what the flaw consists. One could immediately think of a conventional flaw. The provision doubtless offends conventions about the composition of constitutional texts but that in itself does not explain the flaw. For example, a 100-page catalogue of fundamental rights would also be most unusual and unconventional, but despite its unusualness it would not partake in the slightest of what makes the provision about injustice senseless. The same goes when one accepts that there is a moral flaw. From the standpoint of morality it makes no difference if the rights of the majority, at whose denial the provision about injustice aims, are expressly withheld in a second provision. But from the standpoint of what is flawed there is a real difference. The provision about injustice is not merely immoral it is also somehow crazy. One could perhaps think that there is just a political flaw in the provision about injustice.[46] There is such a flaw here, but even that does not explain its flawed nature completely.[47] A constitution can contain much that is politically inappropriate and in this sense technically flawed without it looking as odd as our first provision. Neither the conventional, nor the moral, nor the technical flaws can explain the absurdity of the

[44] This thesis finds a certain parallel in Radbruch's somewhat dark proposition: "Law is that reality whose meaning is to serve the legal value, the idea of law"; see G. Radbruch, *Rechtsphilosophie*, as published in G. Radbruch, *Gesamtausgabe*, in A. Kaufmann (ed.), (Heidelberg: C.F. Müller, 1993), vol. 2, p. 255.

[45] For further examples, see N. MacCormick, "Law, Morality and Positivism" in N. MacCormick and O. Weinberger, *An Institutional Theory of Law* (Dordrecht, Boston, Lancaster, and Tokyo: Reidel, 1986), p. 141; R. Alexy, *Begriff und Geltung*, n. 13 above, p. 68 ff.

[46] E. Bulygin, "Alexy und das Richtigkeitsargument" in A. Aarnio, S.L. Paulson, O. Weinberger, G.H. v. Wright and D. Wyduckel (eds.), *Rechtsnorm und Rechtswirklichkeit. Festschrift für W. Krawietz* (Berlin: Duncker & Humblot, 1993), p. 23 ff.

[47] R. Alexy, "Bulygins Kritik des Richtigkeitsarguments", in E. Garzón Valdés, W. Krawietz, G.H. v. Wright and R. Zimmerling (eds.), *Normative Systems in Legal and Moral Theory. Festschrift for Carlos E Alchourrón and Eugenio Bulygin* (Berlin: Duncker & Humblot, 1997), p. 243 ff.

provision about injustice. This results, as is so often the case with the absurd, from a contradiction. A contradiction comes about because a claim to correctness is necessarily bound up with the act of giving a constitution, and in such cases it is above all a claim to justice. This claim, implicit in the act of giving a constitution, contradicts the explicit content of the provision about injustice. Such contradictions between the content of an act and the necessary presuppositions of its fulfilment can be called "performative contradictions".[48]

The claim to correctness determines the character of law. It excludes understanding law as a mere command of the powerful. Built into the law is an ideal dimension, an "aspiration" in Fuller's sense. This still tells us nothing definitive about Radbruch's formula. But it is clear that the law is not indifferent to its content.

The claim to correctness comprises the eight formal principles which, according to Fuller, define the inner or internal morality of law. But it goes further, including substantive justice[49] and thus what Fuller terms the external morality of law.[50] This connection of formal or procedural aspects with those of a material or substantive kind permits it to take into account the institutional as well as the ideal character of law.[51]

4. The injustice argument

The correctness argument does not by itself suffice to ground Radbruch's formula. While the mere non-fulfilment of the claim to correctness does lead to legal defectiveness, it does not strip a norm or a legal act of its legal character or legal validity. Thus further arguments are required in order to ground Radbruch's formula as a limit of law. The bundle of all these arguments can be called "the injustice argument". It is composed of seven arguments[52] which are in essence normative and which sometimes are made up of several strands.

[48] In this regard, see R. Alexy, "Law and Correctness", n. 37 above, p. 209 ff.

[49] Ibid., p. 214 ff. One can show this to be the case since justice is nothing other than correctness in regard to distribution and commutation and law certainly concerns distribution and commutation.

[50] L. Fuller, *The Morality of Law*, n. 31 above, pp. 44, 96, 132, 224.

[51] See in this regard N. MacCormick, "Natural Law and the Separation of Law and Morals", n. 41 above, p. 114 ff.

[52] This number is not written in stone. Some of the seven arguments can be divided up further which would increase the number. Conversely, they would decrease if one coupled together some of the arguments. In addition, one could simply leave out or add arguments. An example of the latter would be a "semantic argument" which made it the case that for certain purposes a concept of law must be free of morality; see R. Alexy, *Begriff und Geltung*, n. 13 above, p. 72 ff. But this point is already dealt with in dealing with the adequacy of Radbruch's formula from the observer's perspective.

(a) The clarity argument

The first argument to be dealt with is the clarity argument. Hart gave it its classic formulation:

> "For if we adopt Radbruch's view, and with him and the German courts make our protest against evil law in the form of an assertion that certain rules cannot be law because of their moral iniquity, we confuse one of the most powerful, because it is the simplest, forms of moral criticism. If with the Utilitarians we speak plainly, we say that laws may be law but too evil to be obeyed . . .; when we have the ample resources of plain speech we must not present the moral criticism of institutions as propositions of a disputable philosophy".[53]

This objection has a certain force but it is not decisive. A positivist concept of law which rejects the inclusion of any moral elements is indeed *ceteris paribus* simpler than a concept of law which contains moral elements, and simplicity *prima facie* implies clarity. Still it is not the case that every increase in complexity brings a corresponding increase in unclarity. There is little to fear in the way of confusion on the part of jurist or citizen when the formula "Extreme injustice is no law" is built into the concept of law. Confusion could also come about when courts or legal philosophers say to them that the most extreme injustice can be law. It is true that unclarity can come about because of cases like the wall shootings in which the line between extreme and less than extreme injustice is not easy to draw. Still this is not a problem for the clarity argument, only for the legal certainty argument. The clarity argument concerns itself exclusively with the question whether confusion results when moral elements in particular are included in the concept of law.

One should agree with Hart that clarity is a "sovereign virtue in jurisprudence".[54] However one should not agree with his attribution to positivism of the "ample resources of plain speech" and to anti-positivism the "propositions of a disputable philosophy".[55] Anti-positivism can also be formulated in plain speech and positivism can also be seen as a disputable philosophy. In the conflict between positivism and anti-positivism both camps confront each other on fundamentally equal terms. That positivism cannot even claim for itself a presumption of correctness is demonstrated by the fact that the law necessarily raises a claim to correctness. This speaks more for than against the inclusion of certain criteria of correctness in the concept of law. Hence, the clarity argument is not a knockdown one in this respect.

[53] H.L.A. Hart, "Positivism and the Separation of Law and Morals", in H.L.A. Hart, *Essays in Jurisprudence and Philosophy* (Oxford: Clarendon Press, 1983), p. 77 ff.

[54] Ibid, p. 49.

[55] Ibid., p. 78.

(b) The efficacy argument

Radbruch put forward the view that legal positivism had made "both jurists and the people defenceless against just such arbitrary, cruel and criminal statutes".[56] It had "disempowered every capacity to resist National Socialist legislation".[57] His new[58] formula was supposed to provide jurists "with weapons against a recurrence of such an unjust state".[59] In these quotations the future is as much a concern as the past. In respect of the past, we find two theses in Radbruch: a causal thesis and an exoneration thesis.[60] The causal thesis maintains that positivism eased the National Socialist takeover of power in 1933. The exoneration thesis argues that the unjust judgments given by judges of the Third Reich on the basis of unjust laws could "not lead to an attribution of personal responsibility . . . just because they were educated in the spirit of positivism".[61] There are serious objections to both theses,[62] but these will not be pursued here. The acceptability of Radbruch's formula as a thesis of legal philosophy does not depend on Radbruch's conjectures about legal history, whether these are right or wrong. Rather, it depends on whether it, at an altogether general level, contributes somewhat to preventing the worst sort of injustice, thus whether it is effective. This is the future directed aspect of the arming "against a recurrence of such an unjust state".[63]

Hart accused Radbruch of "extraordinary naïvety".[64] One could hardly take seriously that an anti-positivist concept of law "is likely to lead to a stiffening of resistance to evil".[65] The objection about inefficacy is in good part completely justified. It makes little substantive difference to a judge in an unjust state whether he relies on Hart and refuses to apply an extremely unjust law on *moral* grounds or, with Radbruch, does the same by calling on *legal* grounds. In both cases he has to reckon with personal costs and the preparedness to take these on board depends on factors other than the concept of law.

Nevertheless, there are differences from the perspective of efficacy. The first becomes clear when one focuses on legal practice rather than the individual

[56] G. Radbruch, "Fünf Minuten Rechtsphilosophie" in G. Radbruch, *Gesamtausgabe*, (A. Kaufmann (ed.), Heidelberg: C.F. Müller, 1990), vol. 3, p. 78.

[57] G. Radbruch, "Gesetzliches Unrecht", n. 6 above, p. 90.

[58] For the relationship between Radbruch's legal philosophy after 1945 to his (in effect) positivistic stance before 1933, see S.L. Paulson, "Radbruch on Unjust Laws", n. 9 above, p. 489 ff.

[59] G. Radbruch, "Gesetzliches Unrecht", n. 6 above, p. 90.

[60] S.L. Paulson, "Lon L. Fuller, Gustav Radbruch, and the 'Positivist' Theses" (1994) 13 *Law and Philosophy* 313 at 314.

[61] G. Radbruch, "Die Erneuerung des Rechts" in G. Radbruch, *Gesamtausgabe*, in A. Kaufmann (ed.), Heidelberg: C.F. Müller, 1990) vol. 3, p. 108.

[62] See in this regard, S.L. Paulson, "Lon L. Fuller, Gustav Radbruch, and the 'Positivist' Theses", n. 60 above, p. 314 ff. and H. Dreier, "Die Radbruchsche Formel—Erkenntnis oder Bekenntnis?" in H. Mayer (ed.), *Staatsrecht in Theorie und Praxis. Festschrift Robert Walter* (Vienna: Manz, 1991), p. 120 ff.

[63] G. Radbruch, "Gesetzliches Unrecht", n. 6 above, p. 90.

[64] H.L.A. Hart, "Positivism and the Separation of Law and Morals", n. 53 above, p. 74.

[65] H.L.A. Hart, *The Concept of Law*, n. 35 above, p. 210.

judge who measures legalised injustice against his conscience.[66] When there exists in legal practice a consensus that the fulfilment of certain minimal requirements of justice is a necessary condition for the legal character or validity of the rules of the state, then anchored in legal practice is the capacity to provide resistance to the acts of an unjust state by dint of arguments which are juridical as well as moral. In this respect it is true that one should not be under any illusions about the prospects for success of such resistance. A fairly successful unjust regime is in the position quickly to destroy the consensus of legal practice by individual intimidation, changes in personnel, and rewards for conformity. But it is after all thinkable that a weaker unjust regime will not succeed, at least in its beginning phase. This is a relatively limited effect, but still an effect, which we can call the "effect on practice".[67]

Once an unjust state is successfully established, legal concepts can no longer do much work. As the German judicial decisions after 1945 and after 1989 show, they can make a substantive difference only after the collapse of such a state. But somehow there is a delicate and not unimportant effect of the anti-positivist concept of law which can successfully work against legislated wrongs even in a successfully established unjust state. We can call this the "risk effect".

For a judge or official in an unjust state his own situation will look different depending on whether he has reason to interpret it in accordance with a positivistic or an anti-positivistic concept of law. Take for example a judge who confronts the question whether he should impose a terroristic prison sentence which falls within the scope of the legislated injustice. He is neither saint nor hero. He is as little concerned about the fate of the accused as he is greatly concerned by his own. On the basis of historical experience, he cannot exclude the possibility that the unjust state will collapse and he wonders about what would then happen to him. Suppose that he must accept that an anti-positivistic concept of law will prevail or be generally accepted, according to which the norm on which he based his terroristic judgment is not law. It follows that he undertakes a relatively high risk of not being able to justify himself later and thus of being prosecuted. The risk is diminished if he can be sure that his conduct will be judged later in accordance with a positivistic concept of law. The risk does not disappear altogether, because a retroactive law can be enacted on the basis of which he could be deemed responsible, but it is still not equivalent. Given the problems for the rule of law created by retroactive penal statutes it is quite likely that no such law will pass, and if it does pass, he can still defend himself on the basis that he acted in accordance with the positive law of the time. This makes

[66] See W. Ott, "Die Radbruch'sche Formel. Pro und Contra" (1988) 107 *Zeitschrift für Schweizerisches Recht* 335 at 347.

[67] It has been objected that the inclusion of moral elements in the concept of law holds the danger of an "uncritical legitimation" of the law; see H. Kelsen, *Reine Rechtslehre*, n. 12 above, p. 71. Radbruch's formula confines this danger by setting only an outermost limit to law; see R. Alexy, *Begriff und Geltung*, n. 13 above, p. 82 ff. The actual source of this danger is the claim to correctness which the law necessarily raises. But this claim, when taken seriously, provides at the same time the most effective means of combatting the danger.

it clear that a prevalent or general acceptance of an anti-positivist concept of law increases the risks for the individual in an unjust state who goes along with or participates in unjust acts which are covered by statute. It may follow that even for those who see no reason to refrain from participating in injustice, or who would think such participation valuable, an incentive is established or strengthened to refrain from participation in injustice or at least to modify it. In this way, the prevalent or general acceptance of an anti-positivist concept of law can have positive effects even in an unjust state.[68] In sum, one can say that from the perspective of keeping legislated injustice at bay the anti-positivist concept of law in some respects at least has the advantage over the positivist.

(c) The legal certainty argument

A third argument against the anti-positivist concept of law supposes that it endangers legal certainty. In point of fact this argument affects those varieties of anti-positivism which propose a complete coincidence of law and morality and thus say that any injustice leads to the loss of legal character. And when one accords anybody the authority to decide not to follow laws if this is what his judgment about justice requires then the legal certainty argument becomes even stronger—an anarchism argument. We do not have to go further into this, since no anti-positivist who is worth taking seriously has put forward such views.[69] For Radbruch legal security is a value of the highest order. His reference to the "heavy"[70] and "frightful dangers for legal certainty"[71] show that he knew what was at stake. Radbruch's formula is not the result of a natural law intuition or an emotional reaction to National Socialism. Rather, it is the result of a careful balance of three elements which according to Radbruch make up the idea of law, which—as in the case of the claim to correctness—is implicated in the concept of law.[72] The three elements are justice, purposiveness and legal certainty.[73] In 1932 it was the case for judges, though not citizens, that the balance was achieved through giving legal certainty an "unconditional precedence" over justice and purposiveness.[74] In order to get to his famous formula after 1945,

[68] Of course, these positive effects can be accompanied by negative ones. The prospect for the elite of an unjust regime of finding itself in court can strengthen their resistance to the threat of losing dominance. Here obviously a lot depends on the prevailing circumstances. Moreover, it is generally true that dictators and tyrants will only hand over power when there is no other choice and that their helpers and those who do their dirty work will be the more impressed by the risk effect the closer the hour of their downfall.

[69] This may be the reason why Hart speaks of a "danger of anarchy" which older authors like Bentham and Austin "may well have overrated"; H.L.A. Hart, *The Concept of Law*, n. 35 above, p. 211.

[70] G. Radbruch, "Gesetzliches Unrecht", n. 6 above, p. 90.

[71] G. Radbruch, "Die Erneuerung des Rechts", n. 61 above, p. 108.

[72] G. Radbruch, *Rechtsphilosophie*, n. 44 above, p. 255; id., "Gesetzliches Unrecht", n. 6 above, p. 89. See on this F. Salinger, *Radbruchsche Formel und Rechtsstaat* (Heidelberg: C.F. Müller, 1995) p. 7 ff.

[73] G. Radbruch, *Rechtsphilosophie*, n. 44 above, p. 302.

[74] Ibid., p. 315 ff.

Radbruch had to make only a minor adjustment in the system. It establishes a "hierarchy", which corresponds to Radbruch's older positivist understanding, in which purposiveness was at "the bottom" and legal certainty generally preceded justice. Only in the extreme case of intolerable injustice does the relationship reverse.[75] When there exists such a thing as extreme injustice then this way of conceiving the relationship of legal certainty and justice is not only acceptable but mandatory. To give legal certainty precedence even in the case of extreme injustice could not be reconciled with the claim to correctness, which includes justice as well as legal certainty.

(d) The relativism argument

At this stage everything turns on the question whether there is such a thing as extreme injustice. Hart remarked that nothing followed for the concept of law from the fact that moral principles are "rationally defensible" or "discoverable".[76] We shall not attempt to decide this issue here.[77] In any case, the converse is right. If all judgments about justice were nothing more than mere expressions of emotions, decisions, preferences, interests or ideologies, in short, if the thesis of radical relativism and subjectivism were correct, little could be said in favour of the anti-positivist concept of law. Radbruch's formula would then be nothing other than an empowerment of the judge to decide against the law in cases in which his subjective convictions are particularly intensively affected.[78] Hence, anti-positivism presupposes at least a rudimentary non-relativist ethic.

It is not possible to discuss here the problem of the justification of moral judgements or the objectivity of moral knowledge. Radical relativism can be opposed here only by means of a thesis and its illustration using two examples. The thesis states that judgments about extreme injustice are genuine judgments, capable of a rational justification and in so far possessing a cognitive and objective character.[79] Both examples are those already presented—the decisions on National Socialist injustice and on the killings on the internal German border.

The Federal Constitutional Court justified its application of Radbruch's formula in the decision about loss of nationality by saying that:

> "the attempt to destroy physically and materially certain parts of one's own population, including women and children, in accordance with 'racial' criteria"

[75] G. Radbruch, "Gesetzliches Unrecht", n. 6 above, p. 88 ff.

[76] H.L.A. Hart, "Positivism and the Separation of Law and Morals", n. 53 above, p. 84.

[77] Radbruch thought otherwise; see G. Radbruch, *Rechtsphilosophie*, n. 44 above, p. 312: "Doubtless, if the purpose of law and the means towards its achievement could be scientifically and clearly ascertained the conclusion would be inevitable that the validity of positive law must cease to exist which deviates from the natural law which science once recognised, just as the exposed error gives way to the revealed truth. No justification is conceivable of the validity of demonstrably false law".

[78] See on this point N. Hoerster, "Zur Verteidigung des Rechtspositivismus" (1986) 39 *Neue Juristische Wochenschrift* 2480 at 2482.

[79] See R. Alexy, *A Theory of Legal Argumentation* (Oxford: Clarendon Press, 1989), p. 177 ff.

"intolerably" contradicts justice[80] and so amounts to an extreme injustice. This example is decisive. Naturally one can ask the further question why the destruction of parts of one's own population on a racial basis is extreme injustice. This question would however approach what Radbruch called "wilful scepticism".[81] Here we should accept that there is a core area of human rights such that harm to it amounts to extreme injustice.[82]

If this is right then in principle the relativism objection is rebutted. Naturally nothing has yet been said about the degree or scope of the rebuttal. The wall shooting cases show this clearly. In contrast to the destruction of Jews in the Third Reich, there is a serious controversy about whether the killings on the internal German border amount to extreme injustice.[83] The mere fact of this controversy shows that in this case the question whether there was extreme injustice cannot be decided by appeal to evidence but only with the help of arguments. In this regard the issue is not confined to the killing of people on a border. In addition, there is the fact that the killing took place because the fugitive wanted to leave a country in which he had to conduct his whole life in accordance with the will of the political leadership under circumstances which he did not desire and which he apparently detested. Even this might not amount to extreme injustice. But if one considers as a third factor that in the political system which the fugitive wanted to escape there was no possibility of changing the relationships through free public discussion and a political opposition, then there is something to be said for classifying as extreme injustice the killing of for the most part young men at the Berlin wall and on the border strips which until 1989 divided Germany.

Fuller objected that Radbruch's recourse to some or other "higher law" was superfluous.[84] Fuller's target here was what Radbruch defined as "suprapositive law" and specified primarily as human rights.[85] Fuller suggested bringing into play as a substitute for such substantive standards his inner or internal morality of law, that is, his principles of legality:

> "To me there is nothing shocking in saying that a dictatorship which clothes itself with a tinsel of legal form can so far depart from the morality of order, from the inner morality of law itself, that it ceases to be a legal system".[86]

Fuller applied this expressly to "the invalidity of . . . statutes",[87] and, like Radbruch, he worked with a threshold which had to be crossed—"so far depart". One can therefore speak of a Fullerian version of Radbruch's formula.

[80] BVerfGE 23, 98 (106).

[81] G. Radbruch, "Fünf Minuten", n. 56 above, p. 79.

[82] For an attempt to ground such acceptance, see R. Alexy, "Discourse Theory and Human Rights" (1996) 9 *Ratio Juris* 209.

[83] See on this point, R. Alexy, *Mauerschützen*, n. 23 above, p. 23 ff.

[84] L.L. Fuller, "Positivism and Fidelity to Law" (1957/58) 71 *Harvard Law Review* 630 at 659.

[85] G. Radbruch, "Gesetzliches Unrecht", n. 6 above, p. 90; id., "Fünf Minuten", n. 56 above, p. 79.

[86] L.L. Fuller, "Positivism and Fidelity to Law", n. 84 above, 660.

[87] Ibid.

This version has the advantage that Fuller's principles of legality, for example, the requirement of publicity, the prohibition on retroactivity and the requirement of compliance with law[88] offer the relativism objection a much smaller target than Radbruch's substantive standards which are directly oriented to justice. In addition, it is generally true that extreme injustice is bound up with extreme harm to principles of the rule of law. However, the "overlapping" of substantive justice and of the formal requirements of the rule of law which Fuller observed is not strong enough to make Radbruch's formula superfluous. Decree 11 under the Reich's Citizenship Law of 25 November 1941, the topic of both of the cases set out above on National Socialist injustice, was enacted on the basis of an enabling provision[89] contained in section 3 of the Reich's Citizenship Law of 15 September 1935[90] and was published in the appropriate fashion.[91] The Reich's Citizenship Law, which expressly reserved citizenship to those of "German or substantively related blood", thus setting the stage directly for Decree 11, was unanimously approved by the Reichstag. Decree 11 is clear and determinate and it was followed by the official organs of the Third Reich. It does contain certain retroactive elements, because when it came into force it removed the citizenship and property of Jews who had emigrated before it came into force. But that is a relatively weak form of retroactivity. It consists simply in its being coupled to a set of circumstances at a particular point of time and with particular legal consequences, which obtained and persisted in the past. Taken by itself, this does not amount to a nullity.

Finally, nullity results not from the form of the regulation but from its substance, from its extreme injustice. Fuller's criteria can therefore complement but not replace Radbruch's formula. This is true also of section 27, paragraph 2 of the German Democratic Republic's Border Law, which was the focus of the wall killing cases.

(e) The democracy argument

The democracy argument is closely related to the legal certainty and to the relativism argument. It states that lurking in the anti-positivist concept of law is the danger that the judge in answering the call of justice will oppose the decisions of the legislature, which gets its legitimacy from democracy.[92] Since in addition

[88] See L.L. Fuller, *The Morality of Law*, n. 31 above, p. 39.

[89] One can still ask whether the declaration of loss of property in section 3, paragraph 1 of Decree 11 is something which belongs to the domain of citizenship of the state and of the Reich and therefore whether the enabling provision permitted the issuing of that part of the Decree. Here there are arguments both for and against which is perhaps the reason that the Federal Court of Justice left the question open (BGHZ 16, 350 (353)). Section 2 of Decree 11 in contrast is clearly included within the scope of the enabling provision.

[90] RGBl. I p. 1146.

[91] RGBl. I p. 722.

[92] See I Maus, "Die Trennung von Recht und Moral als Begrenzung des Rechts" (1989) 20 *Rechtstheorie* 191 at 193: "The moral argument can thus easily be abused as a substitute for democracy".

this results in an intrusion of the judicial branch into the legislative, the objection can also be formulated as one about the separation of powers. This objection comes up empty when, as in the cases discussed here, the law of dictators is in issue, who know neither democracy nor the separation of powers. But the objection also loses its force at a more general level. Radbruch's formula pertains only to extreme injustice. It works only in a core area. The controls on harms to basic rights exercised by constitutional courts in democratic constitutional states have a content which goes much further. If one wants to present a democracy or separation of powers argument against Radbruch's formula, one must therefore renounce any judicially controlled accountability of the legislature to the basic rights.

(f) The lack of necessity argument

Radbruch,[93] Fuller[94] and Hart[95] agreed that a retroactive law is to be recommended over the application of Radbruch's formula. One could go a step further and say that Radbruch's formula, at least in the period after the collapse of an unjust regime, is unnecessary because the new legislature has the power to override legal injustice by means of a retroactive law. However, this would be no solution if one takes into account the possibility that the new legislature—for whatever reason—is altogether inactive or not sufficiently active. The case discussed above of the Jewish emigrant's deposit of securities shows this with great clarity. If it were left up to the legislature whether she could get restitution of her property, and the legislature remained inactive, she would endure a violation of her rights based on extreme injustice. There are thus cases, required by the claim to correctness, in which Radbruch's formula is necessary to protect fundamental rights. In the actual case, a restitutory statute had been enacted. It provided for a limited period in which demands for restitution could be validated and the emigrant who had returned to Germany had failed to make a timeous claim. The Federal Court of Justice swept this limitation aside with Radbruch's formula and thus prevented the denial through the restitutory law of restitution to the emigrant.[96] This example shows that respect for the rights of the citizen requires Radbruch's formula.

(g) The candour argument

The candour argument asserts that Radbruch's formula leads to a circumvention in criminal cases of the fundamental principle *nulla poena sine lege*. Hart illustrates this argument through the case decided by the Superior Provincial Court Bamberg in 1949 of a woman who wanted to get rid of her husband and

[93] G. Radbruch, "Die Erneuerung des Rechts", n. 61 above, p. 108.
[94] L.L. Fuller, "Positivism and Fidelity to Law", n. 84 above, 661.
[95] H.L.A. Hart, "Positivism and the Separation of Law and Morals", n. 53 above, p. 76.
[96] BGHZ 16, 350 (355 ff.).

thus denounced him to the authorities in 1944 for having made insulting remarks about Hitler. The husband was sentenced to death, but this was not carried out and he was sent to frontline service. The Superior Provincial Court held that, although the conduct of the woman did not violate the law of the Third Reich, it was to be classified as a violation of the law because it "offended against the sense of justice and reasonableness of all right thinking people".[97] It thus convicted her of deprivation of liberty. Hart objected in the following way:

> "There were, of course, two other choices. One was to let the woman go unpunished; one can sympathize with and endorse the view that this might have been a bad thing to do. The other was to face the fact that if the woman were to be punished it must be pursuant to the introduction of a frankly retrospective law and with a full consciousness of what was sacrificed in securing her punishment in this way. Odious as retrospective criminal legislation and punishment may be, to have pursued it openly in this case would at least have had the merits of candour".[98]

The candour argument is the strongest argument against Radbruch's formula but it is not a knockdown one.

The simplest path to its rescue would consist in narrowing its scope of application. One could say that it indeed leads to the conclusion that statutes which justify extreme injustice can never be law or achieve legal validity, but this does not mean that the trust of the actor in positive law should not be protected. The principle *Nulla poena sine lege* must be exclusively connected to this end and must take its bearings solely from enacted and effective norms whatever their content of injustice. The practical significance of Radbruch's formula would then, in order to protect the actor, be limited by the principle *Nulla poena sine lege*.

However, it is better to take the opposite path which consists in a narrowing of the principle *Nulla poena sine lege* by Radbruch's formula. This narrowing is obviously susceptible to limits for two reasons. The first reason is that Radbruch's formula has an exclusively negative character. It does not create new bases of criminality but only destroys particular grounds of justification in an iniquitous regime. The second reason arises out of the distinction between the prescription of the *lex scripta* and the *ius praevium*. Radbruch's formula cannot by definition offend against the prescription of the *ius praevium*—that the act must be punishable before it is undertaken. According to the formula, it is the justificatory ground of an iniquitous regime that is from the outset a nullity. Thus applying Radbruch's formula does not retroactively change the legal situation, it just determines what at the time of the act the legal situation was. Of course from the perspective of the sheer facts of the matter there is a change, and just in this lies the critical bite of Radbruch's formula. This change means that the prescription of the *lex scripta* is not upheld which secured trust in the appropriately enacted and socially effective law which existed at that time. The

[97] OLG Bamberg, Süddeutsche Juristen-Zeitung 1950, column 207.
[98] H.L.A. Hart, "Positivism and the Separation of Law and Morals", n. 53 above, p. 76.

core of Hart's accusation of lack of candour is thus that Radbruch reduced the principle of *Nulla poena sine lege* to the prescription of the *ius praevium* and thus concealed the harm to the prescription of the *lex scripta*. In this way the fact is concealed that there is a choice between "the lesser of two evils". The impression is created that:

> "all the values we cherish ultimately will fit into a single system, that no one of them has to be sacrificed or compromised to accommodate another".[99]

One can in fact abuse Radbruch's formula in this way.[100] The potential of abuse, however, never entails necessity. In Radbruch himself one finds no simulated harmony. The opposite is the case for he talks of antinomies,[101] conflicts[102] and "frightful dangers".[103] Radbruch was clear that his formula involved a choice between two evils and he did not make the slightest attempt to conceal this. That judicial decision-making can take this line is shown especially by the judgment of the Federal Constitutional Court about the wall shootings. Despite some false steps,[104] it is clear that in the end the question is whether it is preferable to incur the cost of a loss in legal certainty or a loss in substantive justice.[105] When it is not diluted by unnecessary extra features, the application of Radbruch's formula cannot be accused of lack of candour.

With this we are at the close of our review of the seven arguments. They showed that many perspectives come into play in the conflict over Radbruch's formula. Most of the objections can be deprived of their force. Against this background, one is weighing the trust of an actor who is active in an unjust state in an enduring justification on the basis of legislated injustice, a basis which supports his deeds, against the rights of the victim and indeed, because of the risk effect, also against the future victim. As a result, everything speaks in favour of not preserving any protection for the trust of the actor, if the threshold of extreme injustice is crossed. Radbruch's formula can thus also be accepted within the domain of criminal law.[106]

[99] Ibid., p. 77.

[100] See on this point, R. Alexy, *Mauerschützen*, n. 23 above, p. 30; id., *Der Beschluß des Bundesverfassungsgerichts zu den Tötungen an der innerdeutschen Grenze vom 24. Oktober 1996* (Hamburg: Vandenhoeck & Ruprecht, 1997), p. 19 ff.

[101] G. Radbruch, *Rechtsphilosophie*, n. 44 above, p. 302.

[102] G. Radbruch, "Gesetzliches Unrecht", n. 6 above, p. 89.

[103] Ibid., p. 90. See on this point, L.L. Fuller, "Positivism and Fidelity to Law", n. 84 above, 655 ff.

[104] See R. Alexy, *Der Beschluß des Bundesverfassungsgerichts*, n. 100 above, p. 19 ff.

[105] BVerfGE 95, 96 (130,133).

[106] Radbruch's formula results in the deed being in violation of the law. The question of individual responsibility, without which the issue of punishment does not arise, is not thereby answered. In its first wall shooting judgment, the Federal Court of Justice described this pressing problem as "very difficult" but then decided simply on the obviousness to the young border soldiers of the injustice because of its extreme character; see BGHSt 39, 1 (34). This conclusion is problematic; see R. Alexy, *Mauerschützen*, n. 23 above, p. 36 ff. The Federal Constitutional Court explicitly contested this conclusion. It reasoned that the extreme character of injustice not always implies its obviousness for everybody; see BVerfGE 95, 96 (142). But the Court then sat on the fence since it allowed the Federal Court of Justice's conclusion to stand that there was subjective evidence simply on the

basis of objective extreme injustice; see R. Alexy, *Der Beschluß des Bundesverfassungsgerichts*, n. 100 above, p. 35 ff. There is something to be said for holding that many young border soldiers, because of their upbringing and their environment, lacked the potential to cultivate the capacity to appreciate clearly the extreme injustice of their act which would be required to confirm their guilt. It would follow that in spite of the violation of the law brought about by Radbruch's formula, they were not just to be punished in a mild way but acquitted; see R. Alexy, *Mauerschützen*, n. 23 above, p. 24 ff., p. 36 ff. Something different is required in the case of their superiors. In the meantime the Federal Court of Justice has decided the case of the shooting of an armed deserter on the Berlin border in this fashion; see BGHSt 42, 356 (362).

3

The Interpretation and Invalidity of Unjust Laws

JULIAN RIVERS[1]

> "Orders are orders, the soldier is told. Law is law, says the lawyer. But whereas for the soldier the duty of obedience ceases when he knows that the order is directed to the commission of a criminal offence, the lawyer—since the last natural lawyer died out about a century ago—knows no such exception to the validity of law and the obedience of subjects to the law. Law is valid simply because it is law, and it is law when in the majority of cases it is capable of enforcement."
>
> Gustav Radbruch, *Five Minutes of Legal Philosophy* (1945)

Introduction

At 3.15 a.m. on 1 December 1984, S, a twenty-year-old East German,[2] placed a ladder against the wall dividing East from West Berlin and started to climb it in an attempt to escape to the West. He had already reached the top of the wall when one of the bullets from the machine-guns of two border guards hit him in the back. He collapsed and was eventually taken to a police hospital where he died at 6.20 a.m. The delay—of more than two hours—before he received medical attention was the result of regulations requiring the suppression of all information about injuries sustained by fugitives; had the man been seen promptly, he would probably have survived.[3]

It has been estimated that 201 people died attempting to cross the internal German border between the erection of the Berlin Wall in the early morning of 13 August 1961 and the fatal shooting of Chris Gueffroy on 6 February 1989. During that time there were more than 4,000 known incidents in which East German soldiers shot to prevent escape.[4] Had the East German state survived,

[1] I am grateful to Trevor Allan, Patrick Capps, Ralf Dreier and David Dyzenhaus for their comments on an earlier version of this essay; also to Tobias Jakubetz for chasing some references. The misinterpretations remain my own.

[2] Strictly speaking East and West Germany are geographical terms, not political ones. One ought to refer to the German Democratic Republic and the German Federal Republic respectively. Since this is cumbersome and a little opaque for some non-German readers, I shall use "East" and "West" instead.

[3] Second border guard judgment, BGH NJW 1993, 141.

[4] Statistics compiled by the Central Information Office in Salzgitter.

no doubt these deaths would simply have joined the endless catalogue of injustice and oppression that is political history. But the East German state did not survive, and after reunification on 3 October 1990 the mills of criminal justice ground into action. Prosecutions were brought not only against the soldiers and officers directly responsible for shootings, but also against members of the East German Defence Council responsible for the entire border regime. While the most prominent of these trials, against the President Erich Honecker, collapsed, owing to his poor health, fellow members of the Council, Albrecht, Strelitz and Kessler were convicted of culpable homicide as secondary parties, receiving prison sentences of between three to seven years' duration.[5] Likewise, several of the ordinary soldiers were convicted and received sentences, in their case often suspended. Not surprisingly, the legality of these convictions was tested in the Federal Court of Justice (Criminal Division), and their constitutionality in the Federal Constitutional Court. Since the shootings at the internal German border were not without basis in the East German legal order, the German courts found themselves for a second time this century considering the interpretation and invalidity of unjust laws.

The Radbruch formula

Theoretical problems concerning the interplay of law and injustice are unavoidably linked in Germany with the name of Gustav Radbruch (1878–1949). He is best known today for a thesis and a formula. The "Radbruch thesis" was the sociological claim that the dominance of legal positivism in pre-war Germany was a significant factor in undermining the resistance of lawyers to Nazi oppression. That thesis is at least controversial, and probably incorrect,[6] but the controversy need not detain us here, except perhaps to emphasise that the Radbruch thesis is sociological, not analytical. Whether positivism *necessarily* inoculates lawyers from critical reflection on the substantive merits of the legal system they operate is not in point. Radbruch's thesis was that, in practice, it did. The formula arose from his definition of law, which was that law is a system of regulations directed towards justice. It was possible that a law might become so unjust as to lose its character as law. Thus the "Radbruch formula" is none other than a modern day version of Augustine's *lex iniusta non est lex*, with the gloss that the *iniusta* must be extreme. This marks Radbruch out as a natural lawyer, but

[5] Landgericht Berlin, 16 September 1993, altered on appeal to Federal Criminal Court, 26 July 1994. See NJW 1994, 2703. Members of the Politbureau (Egon Krenz and others) were also convicted on 25 August 1997. Appeals are pending, but it is unlikely that the legal analysis will change.

[6] Manfred Walther, "Hat der juristische Positivismus die deutschen Juristen im 'Dritten Reich' wehrlos gemacht?" in Ralf Dreier and Wolfgang Sellert (eds.), *Recht und Justiz im "Dritten Reich"* (Frankfurt a. M.: Suhrkamp, 1989). See also Ralf Dreier, "Gustav Radbruch, Hans Kelsen, Carl Schmitt" in *Staat und Recht: Festschrift für Günther Winkler* (Vienna: Springer, 1998), pp. 193–215. An excellent analysis of the Radbruch thesis can be found in Stanley L. Paulson, "Lon. L. Fuller, Gustav Radbruch and the 'Positivist' Theses" (1994) 13 *Law and Philosophy* 313.

contrary to popular belief his experience of the Nazi regime did not trigger a Damascus-road conversion from a strict legal positivism to a full-blooded natural law theory. Undoubtedly, there was a shift in emphasis, but to the end he remained convinced of the practical advantages of a "positivist" account of law.[7]

The most influential of his writings remains a short article he published in the *Süddeutsche Juristenzeitung* in 1946.[8] The key passage—arguably the most-quoted passage in German legal philosophy—ran as follows:

> "The conflict betweeen justice and legal certainty may be resolved in that positive law, secured by command and force, takes precedence even when its content is unjust and unreasonable, assuming however that the positive law does not depart from justice to such an unbearable extent, that it has to give way to justice as 'incorrect law'. It is impossible to draw a sharper line between cases of statutory non-law and law that is still valid in spite of an inappropriate content, but a different boundary line can be drawn with great clarity: where no attempt is even made to achieve justice, where equality, which is the heart of justice, is consciously denied in the creation of positive law, then the law is not merely to be called 'incorrect', it entirely loses its character as law."

The idea that marked Radbruch out as fundamentally a natural lawyer was his view that a focus on positive law serves one moral value, that of legal certainty, and the formula just reproduced was an attempt to identify just when such a procedural value had to give way to the substantive values of proportionality and justice. It is, however, unfortunate that the passage is too often abstacted from the entire article, since the article as a whole makes plain just how important Radbruch thought legal certainty was. The article starts by recounting a number of situations in which German post-war courts were abandoning any attempt to engage in normal legal reasoning and were deciding cases on principles of substantive justice alone. While Radbruch accepted that there were times when this was necessary, he concluded the article with a plea for a return, wherever possible, to traditional legal reasoning, for it was precisely this, as much as substantive justice, that the Nazi terror lacked. Shortly before his death, he emphasised the value of traditional legal techniques when he wrote:

> "The old natural law was not ousted from its hegemony by positivism without reason, and its resurrection brings undeniable dangers. It is all too easy for error and arbitrary power to call itself 'supra-statutory law', thus turning the search for supra-statutory law into a serious danger for legal certainty."[9]

[7] Arthur Kaufmann, "Die Radbruchsche Formel vom gesetzlichen Recht und vom übergesetzlichen Recht in der Diskussion um das im Namen der DDR begangene Unrecht" (1995) 48 *Neue Juristische Wochenschrift* 81 at 82; Stanley L. Paulson, "Radbruch on Unjust Laws: Competing Earlier and Later Views?" (1995) 15 *Oxford Journal of Legal Studies* 489.

[8] "Gesetzliches Unrecht und Übergesetzliches Recht", (1946) *Süddeutsche Juristenzeitung* 105–8; reproduced in the 4th to 9th editions of Gustav Radbruch, *Rechtsphilosophie*.

[9] *Neue Probleme in der Rechtswissenschaft*; quotation reproduced in Arthur Kaufmann, n. 7 above, 85.

While it is easy to see that Radbruch was not a natural lawyer of the classical mould, it is hard to establish his views on the interrelationship between law and morality with precision. In his *Rechtsphilosophie* (1932) he insisted that the concept of law could only be understood in moral terms of certainty (*Rechtssicherheit*), proportionality (*Zweckmäßigkeit*) and justice (*Gerechtigkeit*). But it was not the role of the judge to have regard to anything other than certainty. The shift that Radbruch made post-war was from seeing arguments based on proportionality and justice as purely personal to the citizen (concerning the moral obligation to obey the law), or to the legislator (concerning the duty to make law), to seeing natural law arguments as a ground for judicial "resistance" as well. And this position also found support among a number of other theorists.[10]

However, although the natural law could give the judge sufficient resources to invalidate positive law, it often could not give sufficiently precise guidance on what should replace it. Natural law was thus anarchic in tendency, and the drastic expedient of invalidating a formally correct law was to be strictly circumscribed. Certainly Radbruch did not accept that there were universal and inalienable principles of justice which could be used to fill all the gaps created by the earlier step of invalidation. The replacement law was "supra-statutory", not "supra-positive", to be found expressed in a socially existing widespread legal consciousness. If anything, it was closest to a rather optimistic reading of Anglo-American common law principles.[11] In particular, the natural law by itself could not establish criminality, which always required a positive act of legislation. Since Radbruch also considered that the distinction between a criminal offence and defence was merely technical, the invalidation of a defence on grounds of gross injustice would not automatically revive the "background" criminal offence. There would be a true legal lacuna, since what was once permitted was now totally unregulated. Natural law tended towards an absence of positive law.

The reception of the Radbruch formula by the German courts

In his essay, "Positivism and the separation of law and morals",[12] Hart implied that the Radbruch formula was used by post-war German courts to deal with grossly unjust Nazi laws. In particular, Hart suggested that the conviction of a

[10] Hans Welzel, *Naturrecht und materiale Gerechtigkeit* (Göttingen: Vandenhoeck & Ruprecht, 1951); H. Coing, "Zur Frage der strafrechtlichen Haftung der Richter fuer die Anwendung naterrechtswidriger Gesetze" (1947) *Süddeutsche Juristenzeitung* 61 at 63; see also Hans-Ludwig Schreiber, "Die strafrechtliche Aufarbeitung von staatlich gesteuertem Unrecht" (1995) 107 *Zeitschrift für Strafwissenschaft* 157 at 165; Horst Dreier, "Gustav Radbruch und die Mauerschützen" (1997) 52 *Juristen Zeitung* 421 at 428–9.

[11] There would appear to be a divergence among interpreters of Radbruch whether he did really distinguish suprapositive from suprastatutory principles—and if he did, what it might mean.

[12] (1958) 71 *Harvard Law Review* 593, reproduced in R.M. Dworkin (ed.), *The Philosophy of Law* (Oxford: Oxford University Press, 1977).

woman for informing on her husband, and thus securing his conviction for criticising Hitler's regime, was covert retrospective punishment, unacceptably denying the character of law to formally valid rules of the Nazi legal system. But contrary to Hart's portrayal, as was immediately pointed out in the *Modern Law Review*,[13] the solution of the grudge informer problem had little to do with the Radbruch formula.

The grudge informer case that reached the pages of the *Harvard Law Review*, thus sparking the debate between Hart and Fuller, was one of the earlier decisions in post-war Germany. The court explicitly did not invalidate the Nazi laws which made it a criminal offence to speak disparagingly of Hitler, but argued that since the wife was under no duty to inform on her husband, since she had informed to further her own private motives, and since she must have realised the serious and grossly unjust consequences to which her information would lead, she was guilty of illegal deprivation of liberty, which had always been an offence under the Criminal Code of 1871.[14] This was precisely Radbruch's own suggested solution to grudge informer cases.[15] Although the idea that the criminal process itself could be used as a tool to achieve the malicious ends of a would-be murderer was unusual, it was held not contrary to principle.[16] Of course, one would have to distinguish between cases of the deliberate use of a corrupt judicial system to achieve one's personal ends, and the mere handing over of information for the court to reach its own decision. But such distinctions could, and should, be made. Invalidity was not in issue.

This solution to the grudge informer cases was nevertheless unsatisfactory, because it convicted people for involvement in state oppression regardless of the legality of that oppression at the time. But Radbruch's general approach was consistent with the practice of post-war German courts, which tended to use conventional legal reasoning to establish that crimes had been committed on the terms of the Nazi legal order. In some cases, this was quite straightforward: the courts could deny that certain of Hitler's orders had the correct form of law at all. Thus the secret commands to doctors and hospital administrators to kill patients who were a drain on resources failed to justify the acts of homicide—they were simply outside the formal law-making apparatus.[17] Other cases were more complex and called for a distinction to be drawn between "plausible" and "implausible" interpretations of Nazi law.[18] For example, in the context of grudge informer decisions, where people had used the incitement to disaffection

[13] H.O. Pappe, "On the validity of judicial decisions in the Nazi era" (1960) 23 *Modern Law Review* 260. The original reporting of the decision of the Oberlandesgericht at Bamburg in (1951) 64 *Harvard Law Review* 1005, while not entirely clear, does not state that the court had retrospectively invalidated a law.

[14] H.O.Pappe, n. 13 above, 263.

[15] As set out in *Gesetzliches Unrecht und Übergesetzliches Recht*.

[16] Anglo-American legal systems employ a similar type of argument with respect to blackmail.

[17] OGHSt 1, 321 (5 March 1949).

[18] Monika Frommel, "Die Mauerschützenprozesse—eine unerwartete Aktualität der Radbruch'schen Formel" in *Festschrift für Arthur Kaufmann* (Heidelberg: C.F. Müller, 1993), p. 81 at 89.

legislation to remove personal enemies, the case where a husband was punished for speaking to his wife could not plausibly involve public speech for the purposes of the relevant law,[19] but where a man spoke to his brother and to a work colleague, it could.[20] In that latter case, however, a different flaw was found. The use of the death penalty was held to be unacceptably disproportionate, given that existing law provided that trivial cases of incitement to disaffection (which this was) were to be punished with imprisonment, and only serious cases with the death penalty. Informers would thus be punished where they had deliberately exploited the unjustifiable interpretations of Nazi law by Nazi courts.

Where the question of substantive injustice could not be avoided, prosecution for a crime against humanity or peace under the clearly retrospective Allied Control Council Law 10[21] was a possibility. In content this law largely paralleled the Charter of the International Military Tribunal at Nuremberg and formed the basis for a second stage of denazification both in the twelve trials held before the United States Tribunals at Nuremberg from 1946 to 1949, and in the ordinary German courts.[22] However, the courts were prepared to approve, if necessary, reasoning that relied on inalienable principles of justice, even if they did not refer explicitly to the Radbruch formula. At times, this could go further than Radbrch would have liked in establishing criminality, as for example in a case involving official participation in the transporting of Jews.[23] The Federal Constitutional Court approved the Radbruch formula in principle at an early stage,[24] although at the same time noting its exceptional nature and the need to examine each individual statute to resolve the conflict between certainty and substantive justice. Clear application only came as late as 1968 when the court denied the character of law to the statute depriving Jews who had emigrated to other countries of German nationality.[25] Significantly, the issue in that case was a dispute between relatives as to whether Dutch or German laws of intestate succession should apply, and, in line with Radbruch's own views, not a criminal matter at all.

[19] BGHSt 3, 110 (8 July 1952). Lon Fuller spotted this point, although he makes no reference to the case.

[20] BGHSt 4, 66 (6 November 1952).

[21] Promulgated 20 December 1945.

[22] A brief account can be found in Robert K. Woetzel, *The Nuremberg Trials in International Law* (London: Stevens & Sons, 1960), pp. 219–26. W. Friedman, *The Allied Military Government of Germany* (London: Stevens & Sons, 1947), has a useful chapter on denazification and provides a translation of many of the related procedural directives. Unfortunately the text of Law 10 is not included in the relevant appendix. For an example of the use of Law 10 before the ordinary German courts, see OGHSt 2, 269.

[23] BGHSt 2, 234 (29 January 1952).

[24] BVerfGE 3, 225; 6, 132; 6, 389.

[25] BVerfGE 23, 98. See also *Oppenheimer v Cattermole* [1976] AC 249. The view of the House of Lords as to the effect of article 116 of the German Basic Law was subsequently confirmed in a decision of the Federal Constitutional Court, BVerfGE 54, 53 (15 April 1980).

The legal background to the border shootings[26]

One possible solution to the border guard problem would have avoided tricky questions about the validity of unjust laws.[27] Under section 7 of the Criminal Code, (West) German criminal law applies to acts committed by or against Germans abroad. Although East Germany had been treated as a foreign country by West Germany since the Framework Treaty of 21 December 1972,[28] the Basic Law still defined East Germans without differentiation as German citizens.[29] It was therefore possible to judge former East German soldiers who had shot East German citizens according to West German law. Of course, East German law would still be relevant under the rules of German criminal "conflicts of laws", but it would be limited by public policy considerations, which would filter out offences and defences fundamentally at odds with the values of the German Constitution.[30] However, the majority opinion among academic commentators, which was consistently adopted by the courts, rejected this "international solution". In broad terms, the Treaty of Union 1990 extended Federal (West German) law to the entire territory of the united Germany.[31] One of the modifications to that general principle concerned the application of criminal law in the case of acts committed in East Germany prior to reunification. The preliminary statute to the Criminal Code was amended by the Treaty of Union to require courts not to convict if the accused could not have been punished under East German law valid at the time the act was committed.[32] The international solution would have emptied those provisions of the Treaty of Union 1990 of any significant content.[33] Thus those provisions, coupled with the general requirement that where the law changes between the commission of an act and its prosecution, the milder law should apply, meant that in practice one would have to show that the border guards fell foul of both West and East German law to convict them of any offence.

[26] A short English account of the cases can also be found in Rudolf Geiger, "The German Border Guard cases and International Human Rights" (1998) 9 *European Journal of International Law* 540.

[27] Heiner Wilms and Burkhardt Ziemske, "Gesetzliches Unrecht und Übergesetzliches Recht?" (1994) 27 *Zeitschrift für Rechtspolitik* 170; Joachim Hrushka, "Die Todesschüsse an der Berliner Mauer" (1992) 47 *Juristen Zeitung*, 665.

[28] See BVerfGE 36, 1.

[29] Article 116 of the Basic Law.

[30] This solution was adopted in the occasional earlier prosecutions of East Germans. See (e.g.) OLG Düsseldorf NJW 1979, 59. The relationship between public policy limitations and the ban on retrospective legislation is not clear: see Joachim Renzikowski, "Zur Strafbarkeit des Schußwaffengebrauchs an der innerdeutschen Grenze" (1992) *Neue Justiz* 152 at 154.

[31] Treaty of Union 1990, article 8 and Appendix I C.

[32] Introductory Statute to the Criminal Code, article 315 as amended.

[33] Hans-Ludwig Schreiber, "Die strafrechtliche Aufarbeitung von staatlich gesteuertem Unrecht" (1995) 107 *Zeitschrift für Strafwissenschaft* 157; Klaus Günther, "Anmerkung" (1993) 13 *Strafverteidiger* 18; Knut Amelung, "Strafbarkeit und Mauerschützen" (1993) *Juristische Schulung* 637 at 638; Jörg Polakiewicz, "Verfassungs- und völkerrechtliche Aspekte der strafrechtliche Ahndung des Schußwaffeneinsatzes an der innerdeutschen Grenze" (1992) 19 *Europäische Grundrechtszeitung* 177 at 178.

It is often pointed out that from a practical perspective the choice of route is irrelevant. Whether one considers East German law as foreign or domestic, one is still going to have to ask whether it was so unjust that it should not be applied. But from a theoretical perspective the choice is highly significant, since under the international solution the prosecuting state does not treat foreign law as law, but as potentially mitigating fact, subject to its own conception of justice. Once East German law was treated as part of the domestic system, conflicts between positive law and justice became more pronounced. For present purposes, it meant that the German courts were in the same position as regards East German law as their predecessors half a century ago had been in relation to Nazi law.

There was no doubt that the vast majority of border guard shootings fell foul of West German law; it was the other requirement that proved problematic, for the legal basis of the East German border regime was not completely clear. Soldiers who shot escaping civilians were never prosecuted, and even if they had been, it is unlikely that a judge would have convicted them. The East German government regularly denied in public that the shootings took place, and attempted to cover any awkward evidence. Soldiers involved were bribed to silence with promotion and financial benefits, medical records glossed over the causes of death, for example "heart failure" was recorded instead of "heart muscle damage", bodies were not returned to families and the circumstances of death were not explained. All this would seem to indicate that the shootings were entirely extra-legal, but matters were not that simple.

The attempt to cross the border without official permission was a minor criminal offence under section 213 of the East German Criminal Code. The offence could be aggravated by the use of "dangerous means or methods", which according to the settled practice of the criminal courts, included the attempt to escape in company with others, with a weapon, or with any other object that might assist escape, such as a ladder or grappling irons. Even in the absence of such factors, an escape could be characterised anyway as a "serious disregard for socialist legality". Attempted escapes would generally meet with a prison sentence of more than two years and would therefore be treated as serious offences (*Verbrechen*).

Murder and manslaughter were of course criminal offences as well, for soldiers as much as civilians, but section 27 of the Border Act[34] contained a defence. Since it forms the centrepiece of discussions it is worth reproducing in full:

> (1) The use of arms is the most extreme type of force against persons. Arms may only be used when other physical means with or without other aids remains without success, or obviously holds no hope of success. The use of arms against persons is only

[34] In force from 1 May 1982. Prior to that, the border regime rested on internal military orders. There is some debate about whether the system of internal military orders was consistent with the East German constitution. For both sides of the argument, see Joachim Renzikowski, "Zur Strafbarkeit des Schußwaffengebrauchs an der innerdeutschen Grenze" (1992) 46 *Neue Justiz* 152 and Jörg Polakiewicz, n. 33 above.

permissible if the purpose cannot be attained by the use of weapons against objects or animals.

(2) The use of arms is legitimate, to prevent the immediate commission or continuation of a criminal offence, which under the circumstances amounts to a serious offence (*Verbrechen*). It is also legitimate in the apprehension of persons strongly suspected of having committed a serious offence.

(3) In principle, the use of arms is to be preceded by a shout or the use of a warning shot, unless an imminent danger can only be prevented or removed by targeted use of the weapon.

(4) Arms are not to be used if:

(a) the life or health of non-participants could be endangered,

(b) the persons involved appear to be children, or

(c) the sovereign territory of a neighbouring state would be shot into.

Where possible, arms should not be used against young persons or women.

(5) In the use of arms, the life of persons is where possible to be preserved. First aid is to be given to injured persons having regard to necessary security measures.

As we have just seen, for one reason or another, the vast majority of escapes were treated by the courts as serious offences, and so section 27(2) appeared in general to exculpate the soldiers.

Although the Radbruch formula has an assured place in current German jurisprudence, considerable doubts were expressed whether the Radbruch formula could correctly be used in these cases, or even needed to be. First, the application of the Radbruch formula would recreate a non-existent criminality, which it was not, in Radbruch's view, the function of natural law to do. Secondly, the injustice suffered by the victims of the shootings was clearly not to be compared with the Nazi terror. Fugitives knew that they might be killed and took a calculated risk. While life in East Germany was unpleasant by Western standards, it was not unbearable;[35] as a rule the fugitives were not escaping persecution. So even if the Radbruch formula applied, it was unclear whether the law exculpating the soldiers, while unjust, was so unjust as to forfeit its character as law. Thirdly, it was questionable whether section 27 of the Border Act indeed meant everything the East German government wanted it to mean.[36] The government was not obviously correct to argue that the border regime in all its aspects was covered by the statute. Perhaps some shootings were extra-legal after all. Finally, by invalidating the entire law, which in textual terms was similar to the West German equivalent,[37] one would be striking out a defence for soldiers who had fired in circumstances which in any country would remain free from sanction.

[35] For a valuable English-language account, see Mary Fulbrook, *Anatomy of a Dictatorship* (Oxford: Oxford University Press, 1995).

[36] The East German government always justified the closed border by an appeal to article 12, paragraph 3 of the International Covenant on Civil and Political Rights (1966), which permits states to restrict freedom of movement.

[37] See Knut Amelung, n. 33 above, p. 638.

Applying the Radbruch formula

In spite of these doubts, the first judgment in the East German border guard cases applied the Radbruch formula to convict the accused soldiers.[38] Although the text of the Border Act appeared to be governed by the principle of proportionality, and not dissimilar to equivalent legislation in other Western countries, in practice it was implemented in a way that was grossly unjust. The court pointed out that the education of soldiers was deliberately left vague. The definition of "serious offence", which technically permitted soldiers to shoot at all, was hopelessly unclear, given that a court could find that an escape was serious in the case of practically any potentially successful atttempt. Although the law indicated that an attempt should be made to render fugitives incapable of flight without killing them, in practice, soldiers who killed fugitives were praised. The factual circumstances of the soldiers—the spacing of watch-towers and their automatic weapons—made a proportionate response impossible. For all these reasons, the legal justification for the shootings was apparent only, and used in a way fundamentally at odds with essential and basic requirements of the rule of law. The law, as interpreted and applied at the time of the shootings, was thus invalid and could be ignored, leaving the normal East German criminal law to apply.

Nevertheless, the judgment of the criminal court in Berlin in the first case stands as an isolated example of the explicit and unambiguous adoption of the Radbruch formula to legitimise the conviction of the border guards. Thereafter, the courts took a more nuanced approach. In particular, the approach of the trial court in the second case to be heard, the facts of which were outlined at the opening of this article, contrasted radically with the approach adopted in the first case.[39] This time a different chamber of the *Landgericht* in Berlin considered the wording of the East German Border Act, together with the training the soldiers had received, and concluded that both were oriented towards a principle of proportionality. There was an incremental procedure which the soldiers were to adopt when apprehending people trying to cross over the border: shouting—attempt to reach by foot—warning shot—directed individual fire at legs, if necessary more than once—automatic fire, in any manner, if necessary killing the escapee. Since in this case, the soldiers had used automatic fire straight away, without attempting the less drastic measures first, they were outside the scope of the Act, and so had no lawful defence to a charge of manslaughter.

There was a second strand of reasoning in the first case considered above, which was in many ways similar. The shooting in question had occurred four years later, in early 1989. By that stage, the training given to soldiers had been modified: they were told that their job was to prevent people escaping to the

[38] Landgericht Berlin, 20 January 1992, JZ 1992, 691.

[39] 5 February 1992. The facts and judgment appear in the report of the decision on appeal to the Bundesgerichtshof, 3 November 1992, NJW 1993, 141.

West, but no reference was made to destroying such people if necessary. Rather, the soldiers were simply instructed to shoot at the feet, or to make the person incapable of flight. In practice, however, as we have seen, those who killed escaping East Germans were still commended and rewarded, and there were no disciplinary or criminal consequences for their actions. On the particular night in question, the accused shot the victim after his group captain shouted at him to do so. At his trial, the soldier attempted to argue that he was acting under superior orders, which in certain circumstances would have been a defence under both East and West German law. However, the Court ruled—having, as we have seen, rejected the defence based on the Border Act—that the command to shoot was not an order for the purposes of that defence.

What is noticeable about both these judgments from the *Landgericht* is their resort to conventional—even "legalistic"—arguments. There is an appeal to the fine print of the law to show that the precise requirements for a lawful defence to an otherwise criminal act remained unfulfilled. There are similarities with the reasoning of the post-war courts. Such reasoning easily gives rise to claims that the Court is being disingenuous. Were the judges really implying that had the soldier conformed to the proportional procedure outlined he would have been acquitted? Or would some other technical slip have been identified? Ironically, although the Court in the second trial explicitly eschewed the application of the Radbruch formula, they ended up reproducing the type of conventional legal reasoning Radbruch had commended in criminal cases.

Constructive interpretation

So far, the courts had developed two competing interpretations of the East German Border Act. Under the first, the Act was interpreted according to the manner of its implementation, i.e. it was taken to mean what the East German government wanted it to mean.[40] On this account it was found wanting and invalidated in its entirety. Under the second interpretation, it was taken to mean what the East German government wanted the West to think it meant. It was apparently oriented towards a proportionate response to attempted escapes, and thus some (but not all) of the shootings, being disproportionate, were not covered.[41] One mediating interpretation between these two positions found in the literature was that all the shootings were covered by section 27 except post-arrest killings.[42] The Federal Court of Justice has now issued over a dozen judg-

[40] Supported in the literature by Günther Jakobs, "Untaten des Staates—Unrecht im Staat" (1994) *Goltdammer's Archiv für Strafrecht* 1.

[41] This interpretation is supported by Ralf Dreier, "Gesetzliches Unrecht im SED-Staat?" in *Festschrift für A. Kaufmann*, n. 18 above, p. 57.

[42] Walter Gropp, "Naturrecht oder Rückwirkungsverbot?" (1996) 50 *Neue Justiz* 393; Robert Alexy, *Mauerschützen: Zum Verhältnis von Recht, Moral und Strafbarkeit* (Göttingen: Vandenhoeck & Ruprecht, 1993). See also BGH NJW 1994, 267 in which a post-arrest killing was prosecuted as murder. There was no question in the court's mind that such killings were covered neither by the Border Act nor by military orders.

ments in cases involving border guards, but two are of principal significance for our purposes. The judgment in the first case to be heard (3 November 1992)[43] remains the leading case, but in a later judgment (20 March 1995)[44] the Court took the opportunity to reaffirm its position and answer a number of critics. At first sight, the leading case was confusing, and even contradictory, but subsequent cases saw a shift of emphasis that resolved the confusion. In short, the Federal Court of Justice developed a fourth interpretation of the statute that enabled them to uphold its validity and largely deny its efficacy.

The approach of the Federal Court appeared at first contradictory because it seemed both to deny and to uphold the validity of section 27. It has repeatedly appealed to the Radbruch formula and insisted that section 27 failed to legitimise the behaviour of the border guards "on account of manifest and unbearable contraventions of elementary requirements of justice and human rights". Yet at the same time it has argued that, given the resources of the East German legal system at the time, section 27 could have been interpreted in a way consistent with human rights standards, and if it had been interpreted in this way, the behaviour of particular border guards would usually not have been covered. These two reasons are not presented by the Court as alternatives, but cumulative, which appears inconsistent. The approach also appeared confusing, because as well as appealing to the Radbruch formula, the Court pointed out that East Germany had committed itself to upholding human rights, by binding itself to the International Covenant on Civil and Political Rights (ICCPR), which contains pertinent rights to life and free movement. Surely the Radbruch formula applies quite independently of any state recognition?

The references to East German international treaty obligations led some critics, not without reason, to think that the Court was arguing that the East German legal system had incorporated those international human rights standards into domestic law, and that section 27 was invalid for that reason. References to invalidity by virtue of the Radbruch formula would thus be strictly unnecessary. It then became important to establish whether East Germany had indeed a monist or a dualist system, and since the majority of commentators were fairly clear that the system was dualist, the court was criticised for misunderstanding the East German system. The issue began to get clearer when a case was brought involving shootings before the ICCPR came into force, and the Court was quite content to turn to the Universal Declaration on Human Rights, and East German participation in the United Nations.[45] Not by any stretch of the imagination could this have had internal effect. And by the time the Court gave its second extended judgment, it had become clear that the Radbruch formula was key, and references to formal commitments to human rights standards were secondary.[46] The references to various human rights

[43] BGH NJW 1993, 141.
[44] BGH NJW 1995, 2728.
[45] BGH NJW 1994, 2708.
[46] By the decision of the BGH on 4 March 1996 (NJW 1996, 2042) the tone of the summary has

documents served only to give the Radbruch formula more substance, and references to semi-formal commitment on the part of East Germany served to legitimise the reinterpretation of section 27.

Yet one is still left with the apparent contradiction between an appeal to the Radbruch formula (i.e. section 27 is invalid) and—not "or"—a constructive reinterpretation of section 27 according to human rights standards. Some critics insist that the two strands of reasoning have to be seen as alternative. In the first half, the Court finds that state practice is a plausible interpretation and application of section 27, but one that offends against fundamental principles of justice. In the second half the Court insists that the one proper interpretation of section 27, according to human rights standards, fails to justify the vast majority of the shootings. The point of this constructive reinterpretation, it is claimed, is to show that since section 27 could have been interpreted acceptably under the East German system, it is not retrospective punishment to convict soldiers who relied on the wrong interpretation. However, those critics who wish to choose between the two tend to opt for the first, having the virtue of honesty over the "unrealistic" reinterpretation of East German law.[47]

But the contradiction is indeed apparent, and it is possible to construct a different reason for going on to consider the "human-rights-friendly" interpretation of section 27. The Federal Court of Justice has never said, in so many words, that section 27 is invalid; instead, it states that because of the gross injustice "the defences are irrelevant", or "section 27 . . . in the interpretation signified by state practice can have no justificatory effect". The Court has seen that the Radbruch formula, as applied here, is too crude. If section 27 is valid in the way the East German authorities pretended, gross injustice goes unpunished. Yet if it is invalid in its entirety, the appropriate response to an armed escape, such as might be legitimate in any state respecting the rule of law, would attract criminal penalties.[48] The Court rightly focuses on the injustice of a particular interpretation of section 27, not on the validity of the text itself.[49] This explains why the Court goes on to consider a reinterpretation of section 27 consistent with human rights. For, if a border guard had acted in accordance with such an acceptable reinterpretation, he should not be convicted, since if East Germany had respected the rule of law (which it could have done, but did not),

shifted noticeably in Radbruch's direction: "the command to prevent escape at any price, including the killing of the fugitive, was so grossly unjust that any possible defence under East German law is to be disregarded, because it contravenes legal convictions common to all peoples, directed to human value and dignity. In this application [*sic*!] of East German law the Senate has taken account of international law principles".

[47] Walter Gropp, n. 42 above, at 395; Wilfried Fiedler, "Anmerkung" (1993) 48 *Juristen Zeitung* 206.

[48] In such circumstances one could have argued that the border guard's actions were criminal under (Radbruch-modified) East German law but legitimate under West German law. One cannot blame the court for wishing to avoid this conclusion!

[49] Klaus Günther, n. 33 above, at 23, "natural law in the guise of interpretation"; also Herwig Roggemann, "Zur Strafbarkeit der Mauerschützen" (1993) *Deutsch-Deutsche Rechts-Zeitung* 10 at 12.

the border guard in such a situation would not have been convicted. In short, the reinterpretation of section 27 functions as a concession to the just border guard, not as an alternative means of convicting the unjust border guard.

The approach of the Federal Court of Justice is consistent, but it still fails adequately to reflect the spirit of the Radbruch formula. For Radbruch, it was not sufficient that a law be unjust, it had to be grossly unjust before a court could legitimately refuse to apply it. The effect of the Court's judgment is to criminalise actions that were merely unjust, not grossly unjust in Radbruch's sense.[50] For only if actions could pass muster before a "human rights-friendly" interpretation of section 27 could the soldier be acquitted. If Radbruch's basic idea is to be applied in this area, the approach that matches best, in terms of results as well as methods, is the "legalistic" reading of the trial court in the second judgment. This rejected state practice as a valid interpretation of section 27, but saw section 27 together with the training the soldiers had received as oriented towards a principle of proportionality. Soldiers who observed the gradated response (shout—chase—warning shot—single shot at legs—automatic fire) would be acquitted, but those who reacted hastily, carelessly, or who used the cover of law to satisfy their own violent instincts would be convicted. The human rights interpretation would have convicted the soldier who acted proportionately but who ultimately killed rather than let an unarmed fugitive escape. The issue is not merely whether shooting according to a gradated scheme is unjust; shooting on innocent and unarmed fugitives in any circumstances is unjust. But had it really reached that pitch of injustice that triggered the Radbruch formula?

Reinterpreting retrospectivity

The spectre of unconstitutional retrospectivity hovered over all the border guard cases. A concern that there should be no retrospective convictions had driven the requirement that courts assess the legality of actions under East German law in the first place. While there is no court of appeal beyond the Federal Court of Justice, all persons have the right to challenge the constitutionality of any state action before the Federal Constitutional Court.[51] In the case of such constitutional complaints (*Verfassungsbeschwerden*), the standard of review is the set of rights contained mainly, but not exclusively, in Part I of the Basic Law.[52] A number of defendents issued a complaint, so when the

[50] Gerhard Dannecker and Kristian Stoffers, "Rechtsstaatliche Grenzen für die strafrechtliche Aufarbeitung der Todesschusse an der innerdeutschen Grenze" (1996) 51 *Juristen Zeitung* 490 at 492.

[51] Article 93 I 4a of the German Basic Law. For a brief overview of some of the procedural difficulties, see Julian Rivers, "Stemming the flood of constitutional complaints in Germany" [1994] *Public Law* 553.

[52] Individuals may appeal to one of the fundamental rights contained in articles 1—19, or to the rights contained in articles 20 IV, 33, 38, 101, 103 and 104.

matter came before the Federal Constitutional Court, the key issue was whether the judgments of the criminal courts breached the fundamental right not to be subject to retrospective punishment. The complainants alleged that in refusing to allow them to rely in their defence on the East German Border Act, as that statute was at the time of the alleged offences interpreted and applied in state practice, the criminal courts were punishing them retrospectively. In a judgment of 24 October 1996,[53] the constitutional complaints of three members of the East German National Defence Council[54] and one former border guard against their previous convictions were all rejected.[55]

The complainants appeared to be on strong ground. Article 103 II provides that, "an act may only be punished if its criminal nature is determined by statute before it is carried out". Even if one takes the view that the Radbruch formula is not retrospective, because the higher law invalidates the statute from the beginning, there is still the problem that the criminality of the act must be determined by statute, not just by "law". This is often referred to as the *lex scripta* requirement. The vast majority of commentators saw it as an insuperable obstacle, only resolvable by constitutional amendment.[56] It was argued that article 103 II operated as a requirement of the rule of law quite independently of the justice or injustice of any law. If the German government had entered a reservation in respect of article 7(2) of the European Convention on Human Rights, which waives the ban on retrospectivity in the case of crimes against humanity, how could the lesser offences of East German soldiers not fail to be covered?[57] The Constitution protected the reliance of the soldiers in the continued protection of the law as it was then interpreted, and they could not therefore be held criminally responsible.[58] Exceptionally, Robert Alexy concluded that the reliance of the soldiers in cases of the application of the Radbruch formula was not worth protecting. Article 103 II should be restrictively interpreted in line with its underlying purposes which did not extend to the case at hand.[59] It was precisely the purpose of the formula to create a risk of future punishment for participants in evil regimes.

At first, the Court appeared to be following the vast weight of academic comment. The Court pointed out that the principle of non-retrospectivity was an expression of the rule of law, which itself maintained the basic freedoms by

[53] BVerfG NJW 1997, 929.

[54] Members of the Security Council were convicted of homicide as secondary parties; their convictions thus presupposed the illegality of at least some of the shootings they had instigated.

[55] Further complaints were rejected on 21 July 1997. See Markus Krajewski, "Mauerschützen und Menschenrechte" (1997) 52 *Juristen Zeitung* 1054.

[56] e.g. Kai Ambos, "Nuremberg revisited" (1997) 17 *Strafverteidiger* 39; G. Dannecker and K. Stoffers, n. 50 above; Wanja Welke, "Rückwirkungsverbot zugunsten staatlicher Kriminalität" (1996) 29 *Kritische Justiz* 369 at 375–6.

[57] Horst Dreier, "Gustav Radbruch und die Mauerschützen" (1997) 52 *Juristen Zeitung* 421 at 432.

[58] E. Schmidt-Aßmann, in Maunz-Dürig-Herzog, *Grundgesetz Kommentar*, article 103 no. 255 (Lfg. 30, 1992); Jörg Polakiewicz, n. 33 above, 188.

[59] Robert Alexy, n. 42 above, at pp. 34–5.

ensuring legal certainty, binding state power to law and protecting reliance. Article 103 II itself required the elements of each criminal offence to be adequately formulated in advance, it prevented the courts from imposing a punishment greater than the maximum stated, and it required the criminal law to be regulated with sufficient precision by Parliament, not by any form of delegated legislation. Reinterpretations of the law were as much covered as legislative changes. It also required that any statutory defences available at the time of the act be available at trial, even if they had subsequently been repealed. But it was consistent with the rule of law for defences (unlike offences) to exist at common law, and the issue then becomes to what extent article 103 II protects reliance on the continued existence of such defences. We are now close to the complainants' case, because they were relying on a defence resting partly on statutory authority, partly on state regulation and practice.

However, the Court did not answer this general question, ruling instead that there were circumstances under which the "absolute" protection against retrospectivity was subject to limitations. The strict and absolute protection of reliance presupposed the normal case of an act committed in the Federal Republic under conditions of democracy, the separation of powers and the protection of fundamental rights. The criminal law in such circumstances in broad terms satisfied the requirements of substantive justice. While it was in principle right to apply East German law, that law could not give rise to legitimate reliance where it enacted gross injustice, in particular by criminalising serious wrongdoing (i.e homicide) but then exempting certain acts, allowing soldiers to shoot on innocent and unarmed citizens. Since the reliance of the East German border guards was illegitimate, their convictions were not unconstitutionally retrospective.

The judgment of the Constitutional Court is unsatisfactory in many ways. First, there is the awkwardness of continuing to insist that the ban on retrospectivity is strict and absolute while proceeding to find exceptions. Secondly, the Court failed to consider the arguments of those who suggested that the ban on retrospectivity was developed during the period of absolutism and operates quite independently of the substantive merits of the laws in question. Thirdly, the Court avoided two easier routes out of the difficulty. Like the Federal Court of Justice, it could have argued that reinterpretations are never retrospective. Once one has a requirement that statutes be sufficiently clear, one could legitimately insist that each individual must carry the risk that their understanding might subsequently be shown to deviate from that approved by the courts. However, it was probably correct to assume that reinterpretations can be impermissibly retrospective. Certainly the hypothetical "human-rights-friendly" reinterpretation of section 27 by the Federal Criminal Court was retrospective, if not unconstitutionally so. The other route out of the difficulty was in many ways the most attractive. Towards the end of their judgment, the Court pointed out that the East German government had overlaid the relevant statutory defences with orders to destroy fugitives. This fact was then used by the Court

to support its contention that the complainants had been engaged in gross injustice legitimising their convictions. But if the complainants had systematically instigated soldiers to act beyond the law they could be convicted for simple illegality; no Radbruch formula, no retrospectivity.[60]

In a recent essay, Robert Alexy has suggested a better line of reasoning for the reworking of article 103 II of the Basic Law.[61] The principle of non-retrospectivity should be read as including an implicit limiting clause covering cases of gross injustice.[62] This is an improvement on the Court's reasoning, because the device of implying limiting clauses into apparently absolute fundamental rights is a standard technique of German constitutional law. This is not to deny, however, that hitherto article 103 II has been assumed to be unlimited in its scope.

In general, the reasoning of the Federal Constitutional Court follows the pattern set by the Federal Court of Justice before it: although it claims to be interpreting the written constitution, in fact it too is updating Radbruch. There can really be little doubt that until now the ban on retrospective criminal offences under the German constitution has been treated as truly strict and absolute. On such an account it posed an obstacle to the conviction of the East German politicians and soldiers, in the same way as section 27 represented an obstacle. The validity of the constitutional norm was unquestionable, so it had to be reinterpreted to create the space for their conviction; reinterpreted, that is, according to the requirements of justice. In this process, the Court appeals to examples of formal commitment by the system to those values it sees better instantiated in the new interpretation—just as the Federal Court of Justice appealed to public commitments by East Germany to human rights in its interpretation of section 27. But such appeals are not strictly necessary; or rather, no particular formal commitment by the state is necessary, since the justice of the reinterpretation is its own justification. It remains to be seen whether the European Court of Human Rights will be persuaded to follow the German Constitutional Court in its reinterpretation of retrospectivity.

Hart–Fuller–Radbruch: updating the debate

Regardless of its theoretical sustainability, Hart's defence of the separation of law and morals was inadequate in that it failed to address the question of what the court ought to do in any concrete situation. He presented the matter as a three-way choice: convicting the "guilty" person, acquitting the "guilty" person, and introducing a retrospective law, with a firm preference for the third.[63] Of

[60] It is only fair to point out that procedurally, it was not open to the Constitutional Court to re-evaluate the legal basis for the shootings in such a manner.

[61] Robert Alexy, *Der Beschluß des Bundesverfassungsgerichts zu den Tötungen an der innerdeutschen Grenze vom 24. Oktober 1996* (Göttingen: Vandenhoek & Ruprecht, 1997).

[62] Robert Alexy, n. 61 above, p. 15.

[63] H.L.A. Hart, n. 12 above, at 33.

course, this option is not open to a court, and since it would result in the conviction of the "guilty" person anyway, as Fuller pointed out, it appears simply as a preference for "who should do the dirty work".[64] Given Hart's view that it would have been "a bad thing to do"[65] to let the grudge informer go unpunished, one might suppose that he would have preferred the court to convict, but to acknowledge that it was acting retrospectively. On the other hand, Hart probably thought that on the facts it was worse to convict the grudge informer without a retrospective statute than to let her go free.

The failure to provide a clear solution reflects a more fundamental failure to address the nature of the obligation upon the court when faced with an unjust law. This was Fuller's point in criticising Hart for failing to account for the ideal (and obligation) of fidelity to law.[66] Hart quotes Austin's well-known words about the inconclusiveness of reasoning from the law of God,[67] which may be empirically accurate, but does not tell us if the court is right to ignore such arguments. The clear implication in his criticism of Radbruch for having only "half digested the spiritual message of liberalism"[68] is that a statement of existing law tells the court nothing about what it ought (morally) to do. Radbruch is castigated for the assumption that because something is a law, it ought (morally) to be obeyed. This criticism may have the intellectual virtue of stating the problem clearly, but in failing to relate moral to legal obligation it is practically defective.

Fuller canvassed three possible courses of action for courts faced with the grudge informer laws. They could have interpreted them according to "Nazi principles of interpretation", or they could have interpreted them according to their own standards of interpretation, or they could simply have ruled them invalid. Since adopting Nazi interpretive principles was morally objectionable, and imposing new interpretive principles demonstrated a lack of respect for Nazi law anyway, it was hardly surprising the courts simply invalidated the laws.[69] Fuller was right to suspect that interpretation rather than invalidity lay at the heart of the grudge informer cases, but what he meant by "Nazi interpretive principles" was the Humpty-Dumptyism of making a statute mean whatever you want it to mean, as contrasted with conventional interpretive principles, principles which survived intact in areas not driven by Nazi interests, as they did later on in East Germany.[70] The same fairly crude dichotomy of interpretive technique can be found in his parable of the Purple Shirts and the Grudge Informers.[71] In practice, differences of interpretive technique were not

[64] Lon Fuller, "Positivism and Fidelity to law—a reply to Professor Hart" (1958) 71 *Harvard Law Review* 630 at 649.

[65] Ibid.

[66] Lon Fuller, n. 64 above, 632–3.

[67] H.L.A. Hart, n. 12 above, 30.

[68] H.L.A. Hart, n. 12 above, 31–2.

[69] Lon Fuller, n. 64 above, 655.

[70] As a number of commentators point out, there was such a thing as "socialist legality"; see Hans-Ludwig Schreiber, n. 33 above, at p. 169.

[71] Lon Fuller, *The Morality of Law* (New Haven: Yale University Press, rev. edn. 1969), pp. 245–53, see the arguments of the first and third Deputies. Although in the case of the Speluncean

that extreme, and, as we have seen, the post-war German courts generally went about their task of restoring fidelity to law by the tedious task of assessing each new set of facts against formally valid law, using conventional legal technique. Only rarely did they have to appeal to overriding general principles of substantive justice to achieve the ends they desired.

Like Fuller, but unlike Hart, Radbruch saw moral value in the existence of law (the moral value of legal certainty), and thus could find a way of bridging the gap between legal and moral obligation. Unlike Fuller, however, he did not limit the moral values that informed the law to a set of procedural values, for it is indeed inconceivable that the citizen who participates in the joint enterprise of subjecting human conduct to the governance of rules should care only about the procedures under which those rules are made and interpreted, and not the substantive terms of the enterprise. Abstract justice had a significant, if small, effect on the identification of law.

The interpretive turn in modern jurisprudence requires this well-known debate to be updated in two distinct ways. First, the theories of law at play must be transformed into interpretive accounts, that is, into attempts to portray legal practice in its best light rather than attempts to unearth hidden semantic criteria for the correct use of "law".[72] Because interpretive versions of legal theories incorporate criteria of both fit and justification, they are of necessity morally loaded. At the same time, they incorporate an account of the legitimacy of the legal order, and thus also an account of the legitimacy of coercive acts carried out in the name of that order. Secondly, the focus in the practical question must be shifted to issues of the interpretation of laws, rather than their validity as entire laws. The higher German courts were rightly concerned to establish which interpretation of the relevant statutory and constitutional materials was legitimate, rather than whether the law was valid or invalid in its entirety. As a result the Radbruch formula was recast to strike out unjust interpretations rather than unjust laws. The question is whether the courts gave appropriate weight to the moral principles at play in their exercise of statutory interpretation.

The legitimacy of conventionalism

Ronald Dworkin suggests that legal positivism can be reinterpreted as a theory of legal legitimacy, which he terms conventionalism.[73] To the extent that Hart's concept of law (and adjudication) survives such a transformation[74] it becomes

Explorers ((1949) 62 *Harvard Law Review* 616), Fuller contrasted literal and purposive approaches to statutory interpretation, in general he saw unjust laws as raising issues of validity, not interpretation.

[72] Ronald Dworkin, *Law's Empire* (London: Fontana, 1986), ch. 2.

[73] Ibid., ch. 4.

[74] Hart resisted the move in his "Postscript" to *The Concept of Law*, 2nd edn. (Penelope Bulloch and Joseph Raz (eds.), Oxford: Clarendon Press, 1994), pp. 248–50.

broadly similar to the accounts of Fuller and Radbruch. Conventionalism, for Dworkin, sees law as legitimising governmental force by giving citizens due warning of when force will be applied against them.[75] However, Nigel Simmonds has offered a more plausible explanation for the legitimacy of a "positivist" account of law.[76] He rightly suggests that the prime purpose of law for a conventionalist is not so much to warn citizens of governmental acts, but to provide a set of stable social expectations generally, regardless of the source and content of those expectations. By complete contrast with conventionalism, Ronald Dworkin's own interpretive legal theory sees law as an exercise in substantive political morality. Its aim is to achieve the coherent reconciliation of political virtues such as justice, fairness, and procedural due process.[77] Justice and fairness in this context have distinct and narrow meanings. Justice concerns the decisions that standing political institutions ought to make, considered in the abstract. It determines correct outcomes in the distribution of social goods such as rights, liberties, wealth and opportunities.[78] Fairness concerns the appropriate distribution of political power and the establishment of correct procedures for political decision-taking. It is the characteristic of fair procedures that we trust them to give acceptable outcomes whatever those outcomes turn out to be. In an ideal world there will be no tension between justice and fairness, but in this real world of disagreement and ignorance, respect for the proper allocation of political power can conflict with the requirements of abstract justice. Fair procedures can deliver unjust outcomes and unfair procedures can deliver just outcomes. Integrity, for Dworkin, is the distinct virtue of seeking a reconciliation of justice and fairness based on a set of coherent principles.[79] Law as integrity is legitimate because it relates justice to fairness in a coherent and principled way.

Fuller and Radbruch would have been more amenable to the interpretive transformation than Hart was, and their theories suggest that stability or certainty by itself is indeed inadequate. While many of Fuller's desiderata of law's internal morality clearly serve such interests, they go beyond that to incorporate an element of procedural fairness into law. By contrast, Radbruch takes a slightly different route out of mere stability to incorporate elements of substantive justice. Thus in their different ways, Fuller and Radbruch provide mediating positions between Hart and Dworkin, the one adopting a conventionalism mitigated by considerations of fairness, the other adopting a conventionalism mitigated by considerations of abstract justice. We could call their theories "soft conventionalisms".[80]

[75] R. Dworkin, n. 72 above, p. 117.

[76] See N.E. Simmonds, "Why conventionalism does not collapse into pragmatism" (1990) 49 *Cambridge Law Journal* 63.

[77] R. Dworkin, n. 72 above, pp. 164–5.

[78] References to "abstract" justice are thus not to be read as justice in general, but as the requirements of justice abstracted from the decisions particular institutions have taken about the requirements of justice, which fairness might require one to respect.

[79] R. Dworkin, n. 72 above, pp. 176–8.

[80] Just as "soft positivism" is content to let the identification of law turn partly on moral criteria (see Hart, "Postscript", n. 74 above, pp. 250–4), "soft conventionalism" is used here to indicate

Thus interpretive legal theories see law as providing reasons of political morality that justify public action, and the differences between legal theories turn on what counts as a good legal reason for action. Dworkin's discussions are dominated by a categorisation of these reasons into a triad of justice, fairness and integrity, although at times he also mentions due process. However, the scope of each of these categories is not always clear in his writings. Although when he introduces the concept of fairness, Dworkin adopts a broad definition, in his discussions he tends to limit this to fair legislative procedures, once again revealing the presupposition that law is inextricably bound up with the legitimacy of governmental authority. But there is no reason why fairness should not be more expansively applied to the division of political power generally, along with the establishment of procedures by which such power is exercised. As Fuller clearly saw, the requirements of constancy, generality etc.[81] are procedural limitations of governmental power in the interests of the citizen, as both cooperate in the "enterprise of subjecting human conduct to the governance of rules". And, in spite of Dworkin's hesitations, due process is also an aspect of fairness, allocating power in the context of individual disputes about the appropriate response to a particular fact-situation.[82] The triad of justice (just outcomes), fairness (fair procedures) and integrity (principled coherent solutions) is an adequate categorisation of all possible legal reasons.

Justice and fairness in statutory interpretation

In his earlier work, Dworkin suggested that justice and fairness were reconciled in the process of statute-based adjudication by the requirement that judges should respect the statutory text, emanating as it does from a politically legitimate legislator (fairness), but that where there was a choice of competing interpretations of the text, the judge should select that interpretation which best reflects the abstract moral rights of the parties (justice).[83] T.R.S. Allan has argued that on this earlier account, fairness rightly had no role to play in the process of statutory interpretation.[84] Rather, its function was limited to the simple initial requirement that judges respect statutes, emanating as they do from a politically legitimate legislative procedure. Thereafter, where there is a choice of interpretations, the court should be guided by justice alone.

positions that legitimise law's authority by an appeal not merely to the social order it produces, but to the fact that that order is at least to a minimal extent morally attractive.

[81] Fuller's eight requirements of procedural morality are generality, publicity, prospectivity, intelligibility, consistency, performability, constancy, and congruence between declared rule and administration. See *The Morality of Law*, n. 71 above, p. 39 *et seq*.

[82] See R. Dworkin, n. 72 above, at p. 164. But on p. 338 Dworkin accepts that the ideal of procedural due process can apply to legislation.

[83] *A Matter of Principle* (Harvard: Harvard University Press 1985), p. 16.

[84] T.R.S. Allan, "Justice and Fairness in Law's Empire" (1993) 52 *Cambridge Law Journal* 64.

The German higher courts, in their reworking of the Radburch formula, exemplify this approach to the interpretation of section 27 of the East German Border Act. The Federal Court of Justice had to choose from a range of possible meanings of the statute. The statute could have covered all the shootings, or all of the shootings except deliberate, post-arrest murders, or only "proportionate" shootings, or only "human-rights-friendly" shootings, or indeed no shootings at all—being invalid for some reason. They chose in accordance with justice, which in modern terms, for better or for worse, means in accordance with human rights. Likewise, the Federal Constitutional Court, following Robert Alexy's Dworkinian argument, selected an interpretation of the ban on retroactive criminal liability that it thought accorded better with the requirements of justice.

In this process, the Radbruch formula underwent a subtle modification. Where Radbruch sought to deny the name of law to enactments that were grossly unjust, the revised version seeks to exclude interpretations that are simply unjust. Of course, one could argue that the interest in legal certainty is served by the fact that under the revised version, the choice is between *interpretations*, and not every account of the law is even plausibly an interpretation. No enactment—it would seem—can mean everything or nothing.[85] And yet, when the choice of interpretations on offer covers the range of possible legal responses there is no significant difference. Once I know that a statute might be invalidated for gross injustice, it is irrelevant whether I am convicted because a law I thought was valid is later declared invalid, or because an interpretation I thought correct is later declared incorrect. The bare requirement that the court's analysis of the law must count as an interpretation amounts in itself to an insufficient respect for the interest in legal certainty.

Thus in *Law's Empire*, Dworkin was right to abandon the view that fairness has no role in the process of statutory interpretation.[86] Legal texts do not fall neatly into those that have a clear meaning and those where one must choose from a set of equally possible interpretations.[87] Rather, possible interpretations are more or less plausible, more or less clear, and there will be as much disagreement about when an interpretation is possible as there is about which is ultimately correct.[88] An implausible or unexpected interpretation is unfair in Dworkin's sense because it fails to respect the division of power between legislator, citizen and judge. This is clearly the case where an interpretation deviates from the presumed intention of the legislator; such an interpretation may be more just, but it will be less fair.[89]

[85] Yet some decisions can locate statutory words at the vanishing point of significance: "determination" in *Anisminic v Foreign Compensation Commission* [1969] 2 AC 147; "unlawful" in *R v R* [1992] 1 AC 599 for example.

[86] R. Dworkin, n. 72 above, at p. 338.

[87] Ibid. at pp. 350–4; T.R.S. Allan, n. 84 above, 78.

[88] T.R.S. Allan, *Law, Liberty and Justice* (Oxford: Oxford University Press, 1993), ch. 11 generally and p. 267.

[89] R. Dworkin, n. 72 above, pp. 340–2.

However, it is inadequate to suggest that fairness in statutory interpretation only requires judicial deference to legislative intent. Once again, Dworkin's conception of fairness needs some modification. It is only concerned with giving weight to the one-way projection of legislative power, respecting the statutory texts and their purposes.[90] As Fuller rightly insisted, a statute is an act of communication between government and governed in which both participate.[91] A requirement that laws (and their interpretations) be clear and stable limits government power, effectively reallocating some of it to citizens. The practice of judicial statutory interpretation is a process of replicating and contributing to an ongoing debate about meaning, a debate whose primary participants are governments, legislatures, administrators and citizens.[92] The ordinary meaning[93] is of fundamental significance in a system that respects citizens as legal interpreters. In general, conventional canons of interpretation, which used to be viewed as neutral means of fixing the meaning of a legal text, are better seen as canons of political fairness, to be weighed against considerations of abstract justice.[94] An interpretation that departs from the ordinary meaning of the words used, taken in their context, against the background of existing law and having regard to legislative intent is less fair than one that scores well on such counts.

Radbruch's assumption that gross injustice is needed before the judge can legitimately invalidate a law is correct, since only the avoidance of gross injustice can legitimise the unfairness of striking out a law passed under correct procedures. But the reasoning applies equally strongly in the case of interpretations that depart radically from the interpretations of the creators and addressees of that law. The modification that needs to be made to bring Radbruch into line with the interpretive turn is to abandon the assumption that laws carry their interpretations with them, in abstraction from their interpretive communities. The decision to adopt an interpretation that deviates grossly from that indicated by conventional canons of interpretation, requires a very strong appeal to abstract justice if it is to be legitimate, because it is unfair. This account of statutory interpretation falls between the conventionalist version of legal practice, which effectively requires the courts to track the interpretations of others (primarily legislators, but also citizens), and Dworkin's accounts, which first

[90] It is clear from Dworkin's account that by "statutory purpose" he has in mind the subjective purposes of legislators, not any objective purpose that might be assigned to a statute by an interpreter for reasons of justice.

[91] See *The Morality of Law*, n. 71 above, pp. 207–10; *Anatomy of the Law* (New York: Praeger, 1968), pp. 57–69; see also R. Summers, *Lon L. Fuller* (London: Edward Arnold, 1984), pp. 118–22.

[92] *The Morality of Law*, n. 71 above, p. 91 and pp. 227–32.

[93] The idea of ordinary meaning here is not meant to imply that every word in a natural language has a clearly defined semantic range. It simply assumes the obvious: that language works because there are conventional (and by implication unconventional, and totally novel) ways of using it.

[94] By conventional canons of interpretation I do not include common law presumptions, which are generally statements of justice. Fairness explains why clear language displaces the presumption. The maxim *in dubio pro libertate* balances justice (liberty) with fairness (statutory clarity), for example.

located all interpretive power with the judiciary, and still only grudgingly shares it with the legislator.

The interpretation of unjust laws

In an ideal world, there would be no conflict between justice and fairness, but in the real world the construction of legal meaning has to resolve the tension in each particular case. After all, the prime motive for dispersing political power according to standards of fairness is precisely because we cannot achieve agreement on questions of justice. The ideal of integrity is that working resolutions be consistent across a wide range of cases. As Dworkin indicates, a legal system may become so evil that integrity becomes impossible: no possible interpretation of the community's legal practices is acceptable.[95] Within a moderately evil regime, views will differ as to whether that point has been reached.[96] In one sense, adjudication becomes easier under an evil regime. The requirements of fairness are weakened in the case of laws passed by undemocratic tyrannies, so the radical reinterpretation of such laws in accordance with standards of abstract justice is proportionately more legitimate. But the practical stakes for the conscientious judge whose conception of justice puts him at odds with the government are much higher. Certainly, he does not have to resign in advance, but he is likely to be dismissed, or worse.

Adjudication in the aftermath of an evil regime reverses the difficulties. It is politically easier, but intellectually more complex. Whereas the court under an evil regime is itself a participant in the, admittedly risky, debate about meaning, courts adjudicating in the aftermath of an evil regime are only observers of an historically completed process. They cannot legitimise unfair readings of statutes in the name of integrity by appealing to the consequent reconciliation of justice and fairness within law that their interpretation achieves. The German courts cannot claim to be working East German law pure. This is why retrospective statutes have appealed to many who have considered the issue: in terms of justice, they are no superior, for they achieve the same outcome, but in terms of fairness they go some way to mitigating the procedural unfairness of judicial invalidation.

In East Germany, there was a clear and uniform interpretative practice that treated all shootings as covered by section of the 27 Border Act,[97] and fairness requires that that interpretive practice be given great weight, if not out of respect to the undemocratic legislator, at least to the normal citizen and soldier. The

[95] R. Dworkin, n. 72 above, pp. 101–8.

[96] Compare, in respect of South Africa under apartheid, Raymond Wacks, "Judges and Injustice" (1984) 101 *South African Law Journal* 266 and 295 with David Dyzenhaus, *Hard Cases in Wicked Legal Systems: South African Law in the Perspective of Legal Philosophy* (Oxford: Clarendon Press, 1991).

[97] Robert Alexy, *Mauerschützen*, n. 42 above, pp. 10–14.

interests of fairness are not wholly dissipated by the observation that the legislator was undemocratic. In interpreting section 27 of the East German Border Act, justice and fairness pull in directly opposing directions: fairness (state practice) towards the most expansive reading possible, justice (human rights) towards the most restrictive. In the case of the ordinary soldier, the solution must lie with either the reading requiring a proportionate response, which is arguably a little hard on some soldiers, or with the view that every shooting except deliberate post-arrest murder was covered, which arguably concedes too much. Either way, one is forced to conclude that only in cases of grossly unjust shootings were convictions legitimate. Trigger-happy and malicious soldiers were rightly convicted; others were not. If justice demands the conviction of a wider range of soldiers, fairness suggests that retrospective legislation is the only legitimate way of achieving it.

Whether such a retrospective statute would have required a constitutional amendment depends on one's views as to the role of the Federal Constitutional Court. The new interpretation of article 103 II of the Basic Law, which created the space for the convictions, while it may have been better from the point of view of abstract justice, was extremely unfair, given the strong views of the German interpretive community until now. Even Alexy's later justification, while it can demonstrate a continuity of interpretive technique, cannot avoid the novelty of the conclusion. Although the Court is itself part of the constitutional interpretive community, and is authoritative in the sense of having the last word on the meaning of the Basic Law, one wonders if in this case it did not overrate its own significance. A constitution is the common property of the people who live by it and should be interpreted as such.

Conclusion

The former East German dissident, Bärbel Bohley, once complained that they in the East had asked for justice and got the rule of law. If anything, the German courts have betrayed too much of a desire to achieve abstract justice, and too little concern for the rule of law. They ignored the constitutional allocation of interpretive power presupposed in a modern democracy, acting unfairly in cases not sufficiently legitimised by the requirements of abstract justice. Along with Hart, Fuller and Radbruch one is led inexorably back to the conclusion that apart from certain excessive killings, the better way of securing the convictions of the East German border guards would have been to pass a retrospective criminal statute, along with the necessary constitutional amendment, after a wide-ranging public debate. Of course, the result of such a debate may have been that the law necessary may not have been passed: that is the whole point of having procedures. In broader terms, the German courts have simply joined the fashion for human rights constructivism that has already claimed courts in the USA, Australia and New Zealand, and which looks set to conquer the British legal

systems[98] as well. This constructivism shows itself in an increasing willingness to choose the "best" of a wide range of competing interpretations, according to human rights standards. But this grants all the interpretive power in the state to the judiciary, ignoring the separation of interpretive powers presupposed by conventional methods of interpretation.

Radbruch closed his famous 1946 article with the words:

> "Democracy is certainly a valuable thing, but the rule of law is like daily bread, like water to drink, or the air we breathe. The best thing about democracy is precisely this, that only it is capable of securing the rule of law."

Perhaps it is time for the rule of law to return the compliment.

[98] Under the UK Human Rights Act 1998, section 3, courts will be obliged to interpret statutes in line with the European Convention on Human Rights "so far as it is possible to do so".

4

Legality Without a Constitution: South Africa in the 1980s

RICHARD L. ABEL

Law, like any resource, augments and strengthens existing inequalities of wealth and power. The state uses law to enhance its authority, compel obedience, and levy taxes; capital uses it to organise production and distribution (typically increasing concentration), extract surplus, and evade state regulation. The weak rarely mobilise law and tend to be passive when they are its object. Yet the indispensability of rule-like processes to govern any large organisation also makes law a potential instrument of resistance.[1] Law can be a shield because all states feel some need to follow regular procedures while inflicting punishment. Due process arguments can be asserted against other forms of coercion, such as discharge from employment, eviction from land, expulsion from school, debt collection, censorship, taxation, termination of benefits, and sexual harassment. Law also can be mobilised as a sword to forestall governmental or private action (degrading the environment, for instance) or to demand equality (in *per capita* government expenditures or the weighting of votes).

I. SPEAKING LAW TO POWER

Law distributes political power among regions, between centre and periphery, nation and supranational union, and across ethnic and religious divisions. Law constructs the electoral process: eligibility to vote and ease of voting, districting and voting algorithms (and thus the weighting of votes), qualifications of candidates and term limits, political contributions, access to media, participation by civil servants, the process for qualifying referenda, the role and legality of political parties, corruption and ballot tampering.

Law shapes the legislative process: the number and relation of chambers, the basis of representation, margins for passing ordinary legislation, overriding vetoes, and proposing or enacting constitutional amendments, the legitimacy of

[1] I explore this potential more fully in "Speaking Law to Power: Occasions for Cause Lawyering", in Austin Sarat and Stuart Scheingold (eds.), *Cause Lawyering: Political Commitments and Professional Responsibilities* (New York: Oxford University Press, 1998).

outside activities by legislators (particularly those remunerated by special interests), the role and composition of committees, registration of lobbyists, and investigative powers.

Law constrains the executive more tightly than the legislature. In parliamentary systems, prime ministers stand or fall with their parties' legislative majorities and can be removed by the crown or president. Many countries constrain the participation of executive officials in electoral politics. Law limits the discretion of executive agencies, permitting their actions to be challenged as ultra vires. Because of the potential for abuse, the criminal justice apparatus is particularly tightly regulated. The organisation of the prosecutorial function may influence its behaviour: public employees or privately practising lawyers, appointed or elected, the relationship between investigation and advocacy, special prosecutors for political crimes. Prosecutions may be challenged as politically motivated, tainted by racial or other bias, or under *ex post facto* laws or bills of attainder; alternatively, prosecutors may be accused of insufficient concern for categories of victims (women, racial minorities) or crimes (e.g., police violence). The executive can nullify convictions through pardons or clemency. Law may seek to make the police more representative of community values by requiring them to live locally, or more respectful of "liberal" values by requiring higher educational credentials. Laws create mechanisms for disciplining police, sometimes requiring the participation of civilians, or pitting security forces against each other, or allowing civil remedies. Laws limit police investigative powers: search and seizure, interrogation, admissibility of confessions, wire-tapping and eavesdropping, line-ups, coaching of witnesses, compulsory disclosure of prosecution evidence, grounds for arrest, mandates to inform suspects and witnesses about their rights. Other social control agencies—welfare benefits, tax enforcement, educational institutions, regulatory bodies—are subject to similar legal restrictions.

The judiciary is generally viewed as the last resort for ensuring compliance with the law by legislature and executive (as well as private actors). Courts purport to be less overtly political; they are always open and cannot ignore petitioners; and blindfolded justice claims to be no respecter of persons. Law shapes the structure and thus the behaviour of courts. Those with general jurisdiction are thought to be more independent than those with specialised subject matters. Accusatorial and inquisitorial systems uphold legality in different ways. Some courts exclude lawyers to increase the access and comprehension of lay litigants. All judicial systems have mechanisms for correcting error, although this varies by court, party, and subject matter. Law defines the process by which judges are selected, their qualifications and terms, and how they may be disciplined or removed. In jurisdictions with juries, law determines eligibility, selection, decisional process, and what jurors may hear. Laws attempt to insulate courts from outside influence, limiting media coverage and isolating judges and juries. The capacity of courts to resist political pressure is tested most severely in political trials—prosecutions of those defying state power—and civil cases challenging

fundamental rules or practises—segregation, apportionment, abortion. Courts become production lines for conviction during times of civil unrest, when governments may suspend or curtail rights against self-incrimination, bail, and access to lawyers. When reformers seek to mobililize law proactively, judges are adept at devising reasons for not exercising power: doctrines of ripeness, standing, case and controversy, political question, exhaustion of administrative remedies, and abstention. Many judicial victories are largely symbolic, difficult or impossible to implement. And all can be reversed: by the legislature where parliament is supreme, by constitutional amendment where it is not. Although lay people occasionally mobilise the law themselves, legal assistance usually is indispensable. The demography of the profession may affect access to lawyers and their sympathy and commitment. The size and structure of the profession may affect the cost of legal services. The choice between adversarial and inquisitorial systems, structures of practise, and rules of ethics may affect the balance between fidelity to client and obligations to the legal system and society. Law regulates the non-market redistribution of legal services: state support and the encouragement of philanthropic activity. Government legal aid programmes typically limit the clients served, subject matter of cases, and legal strategies. Case load pressures produce routinisation, replacing vigorous advocacy with reasonable accommodation.

We know, therefore, that law can discipline state and private power in constitutional polities, although its efficacy varies greatly. This chapter asks: under what circumstances can law oppose power *without* a written constitution? I explored that question in one of the most unpropitious settings—the role of law in the last decade of the triumphant struggle against apartheid.

II. SOUTH AFRICAN EXCEPTIONALISM

Law was a central actor in the South African drama because few other weapons were available. Soon after it took power in 1948, the National Party completely disenfranchised blacks. The chambers for Indians and "Coloureds" added by the 1983 constitutional "reform" could be trumped by the white-dominated President's Council. Black voters boycotted elections for local government, which, like the ten homelands, they saw as corrupt puppets. The state drastically restricted extra-parliamentary opposition, outlawing the African National Congress (ANC), the main liberation organisation, and the Pan African Congress (PAC), a rival exclusively black liberation organisation, and detaining thousands under the Internal Security Act. Government banned most public gatherings, censored domestic media, and excluded many foreign reporters, publications and films; it owned all television and most radio stations. The 1985 to 1990 Emergency further restricted political activity, and the government detained 35,000 people. Although Umkhonto we Sizwe (the ANC military arm) increased the number and audacity of attacks, it never seriously threatened the state.

South Africa had long proclaimed fidelity to the rule of law. Government publications boasted of it; legislators praised the independence of the judiciary. Courts occasionally invalidated apartheid legislation, although government invariably overturned these decisions. Courts sometimes acquitted opponents of the regime, as in the 1956–1961 treason trial; but government passed more repressive laws and imprisoned all the opposition leaders or drove them into exile.

South Africa in the 1980s was an extraordinarily inhospitable environment for legal challenges to state power. Parliament was supreme. There was no bill of rights. All judges had been appointed by the National Party, and most strongly supported apartheid. The organised legal profession was supine; opposition lawyers suffered mysterious burglaries, bombings, and assassinations; legal aid was grossly inadequate. Yet the infrastructure for vigorous human rights lawyering began to emerge during this decade with generous funding from Western Europe and North America.

To evaluate the potential of law under such hostile conditions, I examined ten pivotal legal campaigns involving political authority (pass laws, conscientious objection, police torture, state provocation of ANC–Inkatha[2] conflict, censorship, and treason trials), labour (management refusal to negotiate with a militant union and support for a "sweetheart" government-financed Inkatha union), and land (the last forced removal, resistance to incorporation into and independence of a homeland, and opposition to government attempts to eradicate a black township). The victories and defeats illuminate the promise of law under adverse circumstances.[3]

III. PARTY STRATEGIES

Those challenging power must forge and preserve internal unity while exploiting divisions among adversaries. The South African government bribed black leaders with public office and corruption; it fostered enmity between political factions, workers and the unemployed, landlords and tenants, and the African National Congress (ANC) and Inkatha. But it never weaned the masses away from the ANC. Although government assumed the unity of whites, there were clear differences between Afrikaans and English speakers, urban and rural, and large capitalists, petty bourgeois, and working class. Individuals might betray collectivities: the doctor responsible for Steve Biko's[4] death wanted to ensure that the police, not the prisons department that employed the doctor, was held

[2] A black nationalist organisation with a power base among Zulu-speaking South Africans.

[3] For a full account, see Richard Abel, *Politics by Other Means: Law in the Struggle Against Apartheid, 1980–1994* (New York and London: Routledge, 1995); see also the ten individual monographs published by the Centre for Applied Legal Studies, University of the Witwatersrand, Johannesburg.

[4] Biko was at the time of his death the most prominent figure in the black consciousness resistance to apartheid.

responsible for subsequent acts of torture; police and army blamed each other for hit squads.

The South African opposition generally preferred negotiation to confrontation, recognising that law usually favoured the government, which could always rewrite it. Weakness compelled the opposition to adopt an appearance of reasonableness in the hope of persuading officials it could not coerce. Negotiation offered the prospect of a hearing, legal representation, and formal minutes. Government also preferred negotiation, which offered secrecy and control over timing. But agreement, once reached, was costly to repudiate. Courts, too, sometimes negotiated, recognising that coerced judgments were unlikely to be executed. But negotiation presupposed the possibility of compromise. Because the roots of the struggle were indivisible authority, positions were irreconcilable: management would not cede workplace control to labour; conscientious objectors and the military could not agree on criteria, nor could opponents and government settle the boundaries of legitimate resistance.

When negotiation failed, government took executive action: deploying police and army, bulldozing homes, seizing newspapers, banning meetings, closing schools, suspending officials, promulgating regulations, incorporating blacks into homelands. The slower, more public process of legislation risked embarrassment, despite the miniscule Parliamentary opposition. Reasons had to be given. Even an omnipotent government could make mistakes: statutory ambiguities or loopholes, ultra vires executive action. Adjudication was its last choice, with good reason: all three 1980s treason trials collapsed; government could not enforce decrees removing blacks from land. Even legal victories were slow and inefficient. It took too long to evict thousands of residents, convict hundreds of demonstrators, or punish newspapers for each offending word. Protesters and media manipulated the inherent ambiguity of language to voice resistance.

Government was fully prepared to jettison law in favour of force: withholding services to drive blacks off land, detaining thousands without trial, killing and wounding hundreds, and using the military to occupy townships. It wantonly tortured, less to extract information than to humiliate and cow. But most leaders were too committed to be turned by torture, and death made them martyrs. Indeed, the opposition could turn weakness to its advantage. Hunger strikes forced the release of detainees. Prisoners were sanctified and became the focus of campaigns for their release (most notably Nelson Mandela). Rank and file might be incapacitated, frightened or killed, but there were plenty of others to replace them. Indiscriminate state violence transformed arrest, detention and punishment from stigmata into badges of honour. Robben Island occupied much the same position in South Africa that the gulag did in Russia.

Opposition violence played little role in enforcing revolutionary discipline or defiance of the state because there was mass support for boycotts of elections, schools, shops and workplaces, attendance at funerals and demonstrations, and participation in strikes, land occupations, and organisations like the United

Democratic Front (UDF), an alliance of opposition groups, and the Congress of South African Trade Unions (COSATU), as well as the outlawed ANC.

The opposition had to choose not only the mode of engagement but also whether to be proactive or reactive. In political arenas (both in Parliament and outside) proactivity allows control over timing, content, choice of forum and publicity. In courts, which are inherently passive, defence may be the better posture. When the opposition threatened litigation, government offered strategic concessions: paying damages for torture victims or deaths in custody, consenting to injunctions against state violence, postponing removals, offering conscientious objectors non-combatant status. Litigation was more effective as a shield than a sword. After black farmers were forcibly removed from their land, a judgment invalidating that action was worth little (especially once government expropriated the property). But when they infiltrated back onto the land, government could not enforce a new eviction decree. When the 1989 Defiance Campaign integrated parks, beaches, hospitals, and transportation, the regime quietly acquiesced (rather than repeat the mass arrests of four decades earlier). Yet the opposition was unable to secure the punishment of white racists who attacked blacks, security force members who tortured or killed, or black vigilantes who engaged in violence. It might win rights to residence but not the jobs, housing, schools, transportation, or safety necessary to make residence viable. It might save a township from elimination but could not compel reconstruction of houses destroyed or the creation of infrastructure.

Reactive strategies have inherent limits. Those acquitted might be charged anew or detained without trial. A massive campaign saved the Sharpeville Six from the gallows but not from prison. Conscientious objectors secured discretionary (rather than mandatory) sentences but not the end of conscription. The exposure of torture suspended it only temporarily. Blacks could protect their houses from bulldozers but not restore township services.

Confronted by the regime's overwhelming force and political hegemony, the opposition rarely enjoyed the strategic advantages of planning. Most mass action was spontaneous: the 1960s pass law protests, the 1976 Soweto uprising, the 1984 Vaal township unrest. Government, by contrast, could achieve surprise and gain publicity. De Klerk—the last leader of the apartheid regime—was master of the dramatic flourish: releasing Mandela from prison, legalising the ANC, repealing apartheid laws, winning a (white) referendum on his reforms. But government typically used its control over timing to procrastinate, seeking to wear down the opposition and distract media attention.

The government's sense of timing was not perfect, however. It gave a black community ten days' notice of a forced removal, ensuring that the world's media were poised to record the outrage (and compelling the government to postpone the action for three months, by which time the press had lost interest). Government was constrained by proceduralism, whose formality increased with the power exercised—executions representing the extreme. Government controlled only the short run, while the opposition had inexhaustible powers of

endurance. Like other long-oppressed peoples, South African blacks made atrocities the focus of annual remembrances that strengthened their resolve.

Power not only corrupts but also fosters carelessness. Government inadvertently repealed its authority to conduct forcible removals. Its very foundation—apartheid—rendered it ignorant of blacks, leading the state to charge an outspoken *critic* of the ANC with conspiring with the outlawed organisation.

When wielding its legal sword, government sought to define the issues narrowly. It prosecuted as crime what would be legitimate dissent in a liberal polity, using the "common purpose" doctrine to associate all protesters with isolated acts of violence, deploring the "black-on-black violence" that government itself fomented, stigmatising the opposition by prosecuting Winnie Mandela for the abduction and killing of Stompie Mokhetsi. The opposition, by contrast, maintained that apartheid's moral bankruptcy rendered all law political, exposing government action as naked power. Even routine enforcement of criminal statutes was politically motivated—notably pass law prosecutions. When the opposition acquiesced in government efforts to frame issues as legal technicalities—as in challenges to censorship regulations or the excision of territory from a homeland—it predictably lost.

The opposition was strikingly effective in forcing government to engage its broad definition of the issues (assisted by liberal evidentiary rules unencumbered by the needs of juries). The treason trials addressed black aspirations and documented the misery of township life. Objectors exposed military repression at home and atrocities abroad. Black farmers educated courts about their title to land and the imperative of caring about ancestral burial grounds. Judges, in turn, lectured the opposition about democracy, the mandate of social order, and government's good intentions.

Much activity seemed directed toward the "court of public opinion". Daylight is a strong antiseptic for moral as well as physical infection, unmasking claims to legitimate authority as naked coercion. Legal proceedings evaded the blanket prohibition against describing and criticising security forces and prisons. Inquests revealed both Inkatha violence and government complicity, ensuring that victims did not die unnoticed and suspects were named. The presence of the Black Sash—a women's human rights group—prominent clerics, and international observers aborted forced removals. But the effects were ephemeral: Inkatha violence, police torture and removals all resumed when the spotlight shifted and audiences lost interest. Furthermore, South Africa—perhaps inured to being an international pariah—consistently flouted world opinion.

In order to get, hold, and sway audiences, the opposition characterised government actions as atrocities, highlighting the most egregious behaviour, imputing the worst motives. Like American slavery, the pass laws divided families. Prisoners dramatised the hardship of indefinite detention by staging hunger strikes. Blacks forced off their land resembled the Israelites wandering in the desert. Newspapers pictured houses and churches destroyed and graves

untended, focusing on the most vulnerable and innocent—children, mothers, the elderly. But this strategy confronted the limited capacity for moral outrage (analogous to donor fatigue in international relief efforts). Ever more heinous offences must constantly be exposed. And the state also can play this game, alleging terrorism, communism, intimidation, necklacing,[5] people's courts, and "black-on-black violence". Publicity also could backfire, strengthening the intransigence of a government afraid of evincing weakness.

Government used its totalitarian powers to prevent the opposition from being heard. It excluded reporters from forced removals, prohibited them from observing or writing about security forces or prisons, and compelled self-censorship. It alternated between castigating the media and wooing it with trips, tips, and press conferences, which disseminated disinformation. It expelled television from the townships during the States of Emergency. Government officials and Inkatha leaders sued critics for libel, inflicting enormous burdens even on defendants who prevailed.

Finally, government emulated the opposition by claiming the moral high ground through dramatic reform announcements: no more forced removals, repeal of the Black Administration Act, even "the end of apartheid". But promises contain negative pregnants—when did you stop beating your citizens? Once made, they exert pressure for performance. Even when fulfilled, demands escalate. If forced removals have been ended, perhaps the 3.5 million dispossessed should be allowed to return.

IV. THE RULE OF LAW

Most South African judges respected clear statutory language, even when this thwarted the government and their own political preferences. A conservative Afrikaner judge who had imposed harsh sentences on UDF leaders in the Delmas treason trial nevertheless found that a law about homelessness could not be used to regulate those living in "dwellings of a permanent or semi-permanent nature". A court held that Parliament could not meaningfully consent to a forced removal without knowing where the former residents would be sent. Judges sometimes transcended the imperatives of plain meaning, finding that one urban migrant "ordinarily resided" in Cape Town even though she lacked legal accommodation and another had worked "continuously" for ten years although each eleven-month contract ended with his annual month's leave.

Unambiguous facts also created pressure for politically unpopular decisions. Overwhelming evidence of police torture convinced a court to enjoin it. Some judges construed uncertain facts to reach anti-government results. The Alexandra treason trialists were acquitted of all charges despite evidence that

[5] Necklacing involved burning people alive using tyres filled with petrol, which were placed around their necks and set alight.

two had been involved in people's courts (arguably a violation of the law). A young Afrikaner magistrate found that the deaths of three union activists had been caused by crimes and named the Inkatha members responsible. A judge threw out treason charges against union and UDF officials when the defence discredited the state's expert witness and unmasked errors in the transcription and translation of videotapes. Because mass repression requires a bureaucracy to keep detailed records, management documented its union-busting strategy in excruciating detail, and prison doctors recorded all the injuries inflicted by police torture.

Procedural irregularities could frustrate government ambitions. After the two-year Delmas treason trial had ended in lengthy prison sentences of UDF leaders the Appellate Division—South Africa's highest court—reversed because the trial judge had improperly dismissed an assessor. The Supreme Court threw out a management victory in a labour dispute because the Industrial Court judge had been the keynote speaker at a seminar organised by the company's labour relations consultant.

Judges occasionally disregarded the plain meaning of statutes. Although legislative history clearly showed that the Defence Act mandated imprisonment for 1.5 times the outstanding military obligation of those who refused conscription without qualifying as conscientious objectors, the Appellate Division restored full sentencing discretion to judges. A judge effectively nullified the statutory presumption of intent in subversion cases by requiring the state to prove it beyond a reasonable doubt.

The opposition sometimes turned weakness to its advantage. Witnesses for whom English was not the first language took refuge in incomprehension and demanded translation (which gave them more time to frame their answers). Accused invoked the inherent ambiguity of symbols, noting that the black, gold, and green of the banned ANC also were the colours of the government-favoured Inkatha movement. One Alexandra treason trialist refused to identify this grouping of colours as the ANC flag: "the colours are similar to that of the ANC, but there was nothing written, it was only colours". Another said that "by 'force' we mean that there should be strong negotiations"; "fight" meant "we must work honestly and very hard"; "bomb meant the way the workers were going to work". Even the judge helped him out with that last, highly compromising phrase: "he exploded from his starting blocks like Ben Johnson". In acquitting all the accused, this judge acknowledged "I am white. I view it in a certain way. A black person might view it differently".

Yet legality had severe limitations, since the National Party made the rules. A liberal judge (who ultimately resigned from the bench) rejected a community's challenge to incorporation into a homeland because the statute did not require the consent of the community but only that of the jurisdiction from which they were being excised—South Africa! Courts were impatient with technicalities: a statutory interpretation based on the relative meaning of commas and parentheses, an argument that the state had used the wrong law to expropriate land.

Courts construed ambiguous facts as well as laws against the opposition, disregarding evidence of guilt in a murder inquest, refusing to credit overwhelming documentation of management efforts to crush a union, believing whites but not blacks, police but not accused, management but not labour. Some government witnesses showed their contempt for legality by telling stories full of internal contradictions. Police, presumably selected and trained for observation and recall, suffered inexplicable amnesia. Government had no hesitation in using torture to elicit "truth".

Government was at least as effective as the opposition in manipulating procedure. While insisting on Parliamentary supremacy, the (white) House of Assembly displayed hypocritical solicitude for the separation of powers by refusing a petition against incorporation into a homeland because the matter was *sub judice*. A court dismissed a challenge to land expropriation because it named the wrong government official. A homeland parliament repudiated its unanimous decision against accepting independence from South Africa because the matter had not been on the agenda and had passed by acclamation rather than vote. Appellate courts stayed the judgments of lower courts while taking years to hear the matter. Bureaucratic insistence on legal (and illegal) niceties frustrated efforts by blacks to enforce judicially recognised rights to urban residence. Government also created "legal" forms that dispensed with almost all procedural protections: detention without trial, censorship by decree, incorporation into a homeland by fiat. Police and prosecutors failed to investigate or charge crimes against opposition figures; perfunctory inquests failed to attribute responsibility for homicides.

Sometimes judicial review of executive action was an empty form. No sooner was one Emergency Regulation invalidated than the government promulgated another, virtually identical. Although the Home Affairs Minister insisted that "a court can still review any decision I make" suspending a newspaper, the judge hearing the paper's challenge to its suspension conceded that it "was dealing with the opinion of a politician and not a judgement of a court of law", and the court could not "pass an opinion. . .upon the Minister's opinion". After a judge invalidated the incorporation of a community into a homeland, a new law gave the State President unfettered discretion to effect "expedient" incorporations.

Respect for legality tends to be proportional to the visibility of government action. Courts whose judgments would be banner headlines in the press might guarantee urban residence to blacks, but the myriad local officials processing millions of applications refused or failed to endorse their passes. The scrupulous procedural regularity accorded conscientious objectors by the courts coexisted with violent attacks on the End Conscription Campaign by covert security force hit squads. The government backed off from forcibly removing blacks in front of foreign television cameras but did so under cover of darkness three months later. A public inquest into the murder of three labour activists might name the suspected killers, but the Attorney General's refusal to prosecute escaped similar scrutiny. A court might enjoin police torture of prisoners in Port Elizabeth,

but such behaviour persisted in other prisons and soon resumed in the Eastern Cape. Violence and threats intimidated witnesses, severely compromising judicial inquiries; some witnesses were permitted to testify in camera. White magistrates tyrannised over rural blacks, denying permits to meet, withholding pensions and pass book endorsements, and refusing protection against white oppression.

All organisations display loose coupling between apex and base, but South Africa refined this moral division of labour into an art. Judicial recognition of black urban residence rights was frustrated by street-level bureaucrats, who showed far more deliberation than speed (like Southern school districts in the USA). Injunctions and prosecutions against Inkatha leaders failed to curb the violence of followers. Judges consciously remained oblivious of police brutality, and high officials denied knowledge of security force illegalities (and continue to do so before the Commission for Truth and Reconciliation). South Africa created urban councils and homeland governments partly to blame blacks for the illegality, corruption and violence essential to white rule. Hastily recruited, poorly trained black police ("kitskonstabels")[6] repressed the townships. Vigilantes, "black-on-black" violence, and "third force" illegality (security force covert operations) served the same purpose. This racial division of responsibility not only allowed white South Africa to proclaim fidelity to the rule of law but also demonstrated that blacks were unprepared for democracy. But the strategy required that black leaders enjoy real autonomy, which they occasionally turned against the regime, refusing independence, condemning apartheid.

South Africa's pretensions to be a *Rechtsstaat* were built on shaky foundations: Parliamentary supremacy unconstrained by a constitution or judicial review. What does the rule of law mean when government can change that law at will? If a court invalidated a removal, government simply expropriated the land. If a court overturned the declaration of an emergency camp in Oukasie, government disestablished the township. If a court required a homeland's consent to the excision of territory, legislation declared (in Orwellian newspeak) that consent was deemed to have been given and the homeland had *never* contained the territory. The judge conceded: "It is not for me to pass judgment about whether that decision was good or bad".

Yet government did not always deploy its plenary powers. After introducing bills to overturn judgments invalidating a forced removal and a community's incorporation into a homeland, it withdrew both. It retreated from threats to annul decisions extending urban residence and abandoned its appeal from a decree invalidating the declaration that a township was an emergency camp. Overruling judicial decisions consumed more capital than it took to pass ordinary legislation, sometimes more than the regime could afford.

[6] "Instant constables", so called because they were trained in weeks before being sent out armed to deal with the rebellion on the streets.

V. IDEO-LOGIC

Because all domination seeks to transmute raw power into legitimate authority, eliciting consent by offering reasons for actions, the opposition also contested government on the terrain of ideology. The two most dramatic instances were the Moutse community's successful challenge to incorporation into the KwaNdebele homeland and the KwaNdebele women's victorious campaign for the vote. Moutse invoked the grundnorm of apartheid—ethnic homogeneity. Government had reversed other proposed incorporations in order to respect this principle. Moutse also argued that government justified the homelands as vehicles for black self-determination but had refused its request for a referendum. When government advanced geographic contiguity and administrative convenience to support incorporation, Moutse argued that the regime had disregarded these factors in constructing other homelands.

The role of ideology was more critical, and paradoxical, in the women's vote case. Whereas Moutse could invoke statutes and administrative law, KwaNdebele women had to rely exclusively on analogy and natural law. Furthermore, they were in the anomalous position of challenging *sex* discrimination in the only country whose official policy was *race* discrimination. They pointed to the illogic of disenfranchisement: Ndebele women could vote outside KwaNdebele; Moutse women could not vote within it. When government claimed that African tradition denied women equality, plaintiffs responded that Ndbele women played a larger role in traditional tribal councils than in the homeland's "modern" legislative assembly. Furthermore, "traditional" inferiority did not prevent them from voting in every other homeland and every urban council. When government sought to justify the homelands as a means of introducing "Western" political institutions, KwaNdebele women responded that apartheid separated men from their families, forcing women to manage the farm, household, children's education, and locality; they deserved an equal political voice.

Resistance to KwaNdebele independence also exposed the delusions of grand apartheid. KwaNdebele residents braved death—not in pursuit of national independence but to remain South African subjects! Homeland governments were supposed to be more democratic and enlightened than "traditional" tribal leaders, but the Ndebele royal house was more progressive than the tyrannical, corrupt homeland government. The Chief Minister, whose only claim to legitimacy was "modernity", had almost no formal education. Once KwaNdebele decisively rejected independence, the foundation of grand apartheid disintegrated (since it was clear that none of the other five homelands would accept it).

Warrants of modernity and tradition also suffused Magopa resistance to forced removal. The leader of the group opposing it advanced the traditional claim to be a "chief by the people", while disparaging the man favouring removal as a government puppet who owed his title to the whim of a white mag-

istrate. The Magopa maintained that traditional farming techniques limited the area to which they could be relocated; government retorted that younger farmers could learn new methods. The Magopa declared that tradition obligated them to maintain ancestral graves; the state's (white) expert called their cemetery a modern innovation.

The opposition sought to heighten apartheid's fundamental contradiction: whites demanded political domination and social separation but depended on black labour for production and domestic service. Large corporations disliked pass laws, which divided worker loyalties between urban employment and rural families. Township residents enlisted their white employers in opposing the removal advocated by other white residents. Blacks opposing Moutse's incorporation into KwaNdebele appealed to neighbouring white farmers. Some local white Afrikaners wanted the Magopa's labour and purchasing power. The opposition's ideological leverage was enhanced by the government's desire to appear reformist in order to appease both internal and external critics. Removals could no longer be justified by white prejudice but only by black betterment. But if they were truly voluntary, residents could obstruct them by refusing to leave. And if the goal was urban development, residents could show that upgrading was cheaper than removal.

The Alexandra treason trial turned into a contest for the mantle of democracy. Those accused of subversion could point to the facts that they kept detailed minutes, sought legal advice, were obsessed with drafting a constitution, held public meetings, were accountable to members, repudiated fiscal irregularities, and condemned adventurist behaviour. Tainted by its racist national franchise, the state pointed to the excesses of people's courts, election boycotts, and the failure of "Coloureds" to appeal to their House of Representatives.[7] The defence won this contest when the judge made the unprecedented admission: "While white South African citizens may have a democracy, Black South African citizens certainly have no share in it".

When conscientious objectors advanced religious and ethical critiques of conscription, the military, prosecutors and judges felt compelled to reply in kind, invoking their Christian faith, the "total onslaught", atheistic communism, the utopianism of pacifism, and the objectors' inconsistent beliefs and flawed characters. Revealing its guilty conscience, government sought to justify censorship as scientific, necessary, the only alternative to media autarchy, while also maintaining that South Africa was the freest country on the continent.

The opposition could draw strength from ideology only if government cared about the coherence and integrity of its own justifications. But the regime often was content with vague, ambiguous reasons. One of the most manipulable was "choice": blacks *chose* to leave their farms or townships; the KwaNdebele Legislative Assembly *chose* independence; conscientious objectors *chose* jail. In

[7] A puppet chamber in the so called "Tri-Cameral Parliament": an attempt by the National government to ward off demands for democracy by giving "Coloureds" and "Indians" each a nominal representative body.

other examples of Orwellian newspeak, the Natives (Abolition of Passes and Coordination of Documents) Act required all Africans to carry registration books; the Department of Co-operation and Development coerced blacks and subjected them to underdevelopment; and the Civil Cooperation Bureau bombed, tortured, kidnapped, and assassinated. Some rhetoric was simply ludicrous. Allowing forcibly removed blacks to return home would endanger the rights of those who chose exile—this from a government that had ruthlessly uprooted millions of blacks over four centuries. The courageous return of other blacks to ancestral land in defiance of the state was a "forced removal" because charity had defrayed transportation costs. Tired of giving reasons, government resorted to the exasperated parent's *ipse dixit*. It could not stop a removal *it* had ordered and was bound by treaties *it* had imposed on homelands *it* had created. It claimed that the irrevocability of decisions—like the value of wine—increased with time. P.W. Botha—next to last leader of the apartheid regime—presented himself as national patriarch: "I'm a hard man; you should know better than to ask me to change my mind".

VI. THE LESSONS OF SOUTH AFRICA

The uniqueness of South Africa makes generalisation dangerous. A small minority oppressed the vast majority; the world had long repudiated the regime's racist ideology; it could claim few supporters abroad, most of them morally compromised (Taiwan, Israel). Extrapolation to post-apartheid South Africa is equally risky, since much depends on the transformation of the state bureaucracy and rates of economic growth.

But the struggle against apartheid offers lessons for resistance to other oppressive regimes. Law was by no means the only or even the most important factor in the ultimate victory. The 1980s also saw the revival of anti-apartheid organisations destroyed by the post-Soweto repression—most notably the UDF. Global competition forced large capital to raise productivity by fostering collective bargaining as an antidote to labour unrest; but the more militant black trade unions united in Cosatu, further pressuring government. Economic sanctions (especially the calling of international loans), amplified by the global recession, intensified fiscal anxiety; the cultural boycott—especially of sports—profoundly hurt national pride. The contradiction deepened between large capital, eager to relax apartheid, and white working class, civil servant, farmer, and petty bourgeois voters fearful of losing jobs, political hegemony, social status, and land. With the end of the Cold War, South Africa could no longer invoke the threat of communism (the "*rooi gevaar*").[8] The cost of military operations—in both economic terms and white boys in "body bags"—pushed the front line back into South Africa.

[8] Meaning "red danger".

But the recognition that law was only one ingredient in the struggle should not diminish its value. The opposition was able to negotiate agreements, which the regime then found it difficult to disavow. When the opposition took direct action, the regime sometimes backed away from confrontation and repression. Even an omnipotent government made errors: setting itself deadlines, repealing legislation under which it sought to act. The opposition succeeded in framing issues broadly, exposing the politics underlying ostensibly legal actions. The opposition effectively used publicity to embarrass the government. South African judges generally enforced clear legal restraints on government power and accepted unambiguous facts supporting opposition arguments. They insisted on procedural regularity. Sometimes they even interpreted ambiguous laws against the government. And though government had the power to overturn any judicial decision, it occasionally refrained from doing so. Finally, the coherence of ideology placed limits on government action.

Hope is essential to resistance; its importance increases as the odds worsen. Opposition legal victories demonstrated the regime's vulnerability and eroded its will to repress. They empowered the masses while offering activists protection from state retaliation. They strengthened the opposition's own commitment to legality and thus the prospect that the post-apartheid regime would respect the rule of law. This quixotic victory continues to offer a beacon to other struggles.

5

Delivering Positivism from Evil

ANTON FAGAN

INTRODUCTION

For several decades, though thankfully no longer, South Africa provided legal philosophers with a clear and uncontroversial example of a wicked legal system.[1] Yet, while legal philosophers were united as regards the wickedness of the South African legal system, they were divided about the import of this for legal philosophy. For some, the wickedness of the South African legal system provided a telling counter-example to Ronald Dworkin's theory of law and adjudication.[2] For others, typically legal academics closely connected with South Africa, the wickedness of the South African legal system supported an altogether different conclusion. To wit, the invalidity of legal positivism.[3]

In this essay I offer a response to the second of these ideas. I explain why the South African legal system's wickedness constituted no threat to legal positivism. In the course of this explanation I will be locking horns with one foe in particular, namely David Dyzenhaus, the editor of this collection of essays. For Dyzenhaus's *Hard Cases in Wicked Legal Systems* represents faraway the most lucid and thorough attempt to found a case against legal positivism upon the wickedness of the South African legal system.[4]

Reduced to its bare bones, Dyzenhaus's critique of positivism is as follows. One: acceptance by a judge of the positivist analysis of law results in acceptance also of the "plain fact" approach to law and adjudication. Two: judicial acceptance of the plain fact approach has undesirable moral consequences. Three: if

[1] See for example Ronald Dworkin, *Taking Rights Seriously*, 2nd imp. (London: Duckworth, 1978) at p. 326; Lon Fuller, *The Morality of Law*, rev. edn. (New Haven: Yale University Press, 1969) at p. 160; H.L.A. Hart, *The Concept of Law*, 2nd edn. (Oxford: Clarendon Press, 1994) at p. 200, *Essays on Bentham* (Oxford: Clarendon Press, 1982) at p. 150; Joseph Raz, "Dworkin: A New Link in the Chain" (1986) 74 *California LR* 1103 at 1111.

[2] See Hart, *Essays on Bentham*, n. 1 above, at p. 150; Raz, "Dworkin: A New Link in the Chain", n. 1 above, at 1111. But compare Etienne Mureinik, "Dworkin and Apartheid" in Hugh Corder (ed.), *Essays on Law and Social Practice in South Africa* (Cape Town: Juta, 1988) at pp. 206–9.

[3] See for example John Dugard, "The Judicial Process, Positivism and Civil Liberty" (1971) 88 *South African LJ* 181. But, for a partial retraction, see John Dugard, "Some Realism about the Judicial Process and Positivism—A Reply" (1981) 98 *South African LJ* 372.

[4] David Dyzenhaus, *Hard Cases in Wicked Legal Systems* (Oxford: Clarendon Press, 1991).

judicial acceptance of an analysis of law has undesirable moral consequences, the analysis is invalid. Four: thus the positivist analysis of law is invalid.

Of the three premises in Dyzenhaus's critique, it is to the first that greatest attention will be paid in this essay. In the section that immediately follows this introduction I explain just what Dyzenhaus means by the "plain fact approach". And I show why, according to Dyzenhaus, acceptance of the plain fact approach follows from acceptance of legal positivism. As we shall discover, Dyzenhaus believes that the connection between positivism and the plain fact approach is forged by the positivist rationale for the sources test. I disagree. In order to defend my view against Dyzenhaus's contrary one I attempt, in section II of this essay, to set out what I take to be the best account of the positivist rationale for the sources test. In the section thereafter I explain why, so understood, the positivist rationale does not have the effect that Dyzenhaus claims for it. That is, I explain why the positivist rationale does not, after all, bridge the gap between legal positivism and the plain fact approach.

The greater part of Dyzenhaus's *Hard Cases in Wicked Legal Systems* is aimed at substantiating the second of his critique's premises. Dyzenhaus provides a detailed historical account of the South African judiciary's record during the decades of apartheid. From this case study, Dyzenhaus draws a twofold conclusion. One: South African judges during this period by and large followed the plain fact approach. Two: their doing so had undesirable moral consequences. Somewhat crudely put, because of their adherence to the plain fact approach, South African judges did less to contain the wickedness of apartheid statutes than they could (or even would) have done, had they adopted a "common law" approach to adjudication more in keeping with Ronald Dworkin's "law as integrity". Of course, Dyzenhaus's critique of legal positivism would make itself vulnerable to a charge of parochialism if the plain fact approach had undesirable consequences only in a wicked legal system like that of apartheid South Africa. Presumably conscious of this pitfall, Dyzenhaus devotes a chapter in his book to a second case study. Its focus is the English judiciary's record in dealing with security matters. According to Dyzenhaus, this case study shows "that the lesson to be learned from . . . wicked legal systems is also one for benign legal systems".[5] For, claims Dyzenhaus, when adjudicating issues of national security, the English courts have also employed the plain fact approach. And, as in the South African context, this has had morally undesirable results.

Some people may be unhappy with the way that Dyzenhaus seeks to support his second premise. They may deny that the plain fact approach was adopted by the South African judiciary during the apartheid years and by the English judiciary when dealing with security matters. Or they may argue that, even if Dyzenhaus is right in attributing acceptance of the plain fact approach in these two instances, he is wrong in concluding that this acceptance had morally undesirable results. In this essay, however, I will not be doing battle on either of these

[5] Dyzenhaus, n. 4 above, at p. 209.

two fronts. For, even if Dyzenhaus were mistaken on both these counts, he would be right to regard the plain fact approach as unacceptable. As we shall see in the next section, the plain fact approach is made up of a rather curious assemblage of jurisprudential ideas. So odd is this assemblage that no legal philosopher has ever defended it. Nor is any legal philosopher likely to do so. Indeed, from a jurisprudential perspective the plain fact approach strikes one as so obviously flawed that one is left wondering why Dyzenhaus felt it necessary to devote so much of his book to demonstrating its inadequacy.

Even more puzzling than Dyzenhaus's loquacity about the second premise in his critique is his virtual silence about the third. Dyzenhaus leaves us in no doubt as to his approval of the third premise. "We should . . . adopt the view of law that gives us the best results in practice", he claims.[6] And he says: "[W]hat more could be asked of the correct view of law than that it lead to morally good results and that it make sense of and perpetuate healthy legal practice?"[7] But Dyzenhaus makes no attempt to provide a defence of the third premise. Presumably he regards its truth to be so self-evident as to be beyond dispute. In the conclusion to this essay I will suggest that the third premise is not only contested, but also false.

Although, for the reasons to be explored in this essay, Dyzenhaus's critique of legal positivism fails, it contains an important insight, namely: acceptance of the positivists' sources test compels acceptance of further criteria for law. A judge who endorses the sources test must also endorse the positivist rationale for the sources test. The positivist rationale, however, does not merely justify the sources test. It also generates other criteria for law. In section IV of this essay I explain why this is so. And I describe two of the criteria for law that are generated by the positivist rationale. They are, first, that law must possess certain formal characteristics and, secondly, that law must emanate from institutions which possess certain formal characteristics. As I further explain in section IV, once legal positivism is revised so as to accommodate these formal criteria in its test for law, it might be that even a positivist judge could have regarded as non-law some of the directives issued by the South African Parliament and government during the apartheid era.

In section V I finally come to discuss the rule of law. As we shall see, some critics of the South African judiciary's record under apartheid assert both that the rule of law has a substantive (as opposed to merely a formal) content and that non-compliance with the rule of law constitutes a legal (rather than merely a moral) failure. This combination of views allows for a powerful condemnation of the South African judiciary's role under apartheid. It also, so these critics appear to believe, adds a second string to the anti-positivist's bow. Because legal positivism cannot endorse these two views, it must be unsound. I shall explain that these critics are right about the incompatibility of their understanding of the rule of law and legal positivism. A legal positivist can give the rule of law a

[6] Ibid., p. 269.
[7] Ibid., p. 270.

substantive content or a legal positivist can treat the rule of law as a legal doctrine. But a legal positivist cannot do both these things in conjunction. However, I shall question the critics' assumption that this incompatibility counts against legal positivism. Might it not as well count against the critics' understanding of the rule of law?

I. THE PLAIN FACT APPROACH AND LEGAL POSITIVISM

The "plain fact" epithet will be familiar to anyone acquainted with Ronald Dworkin's *Law's Empire*.[8] Yet, such familiarity is more likely to impede than to facilitate an understanding of Dyzenhaus's critique of positivism. For, although Dyzenhaus's use of the expression "plain fact" has something in common with Dworkin's, there are important differences. The plain fact view of law, as Dworkin describes it, makes two claims. One is that "law is always a matter of historical fact and never depends on morality"[9] or, more specifically, that "law is only a matter of what legal institutions, like legislatures and city councils and courts, have decided in the past".[10] The other is that, as regards some legal issues, "[t]he law may be silent . . . because no past institutional decision speaks to it either way".[11] Hence, "the judge has no option but to exercise a discretion to make new law by filling gaps where the law is silent and making it more precise where it is vague".[12]

The plain fact approach, as Dyzenhaus sets it out, reiterates the first of the claims identified by Dworkin. The plain fact approach, says Dyzenhaus, assumes "that law is a particular kind of social fact – law is what meets a particular sources test".[13] The second of the claims identified by Dworkin is, however, directly contradicted by Dyzenhaus's plain fact approach. Dyzenhaus writes that "there is no suggestion in the plain fact judgments nor in the picture of the approach itself that the judges thought they had a discretionary power".[14] Moreover, Dyzenhaus's plain fact approach adds two further claims all of its own. One is that the law of a community is to be identified so as to reflect the intentions of the law-makers.[15] The other is that the law has legitimate authority. That is, whenever the law provides a solution to a case, a court is bound to follow that solution regardless of the court's view on its moral merits.[16]

The differences between Dworkin's plain fact view of law and Dyzenhaus's plain fact approach are significant. Remember that Dyzenhaus seeks to link the

[8] Ronald Dworkin, *Law's Empire* (London: Fontana Press, 1986).
[9] Ibid., p. 9.
[10] Ibid., p. 7.
[11] Ibid., p. 8.
[12] Ibid., p. 9.
[13] Dyzenhaus, n. 4 above, at p. 211.
[14] Ibid., p. 219.
[15] Ibid., p. 57, pp. 217–18.
[16] Ibid., p. 57, pp. 217–18.

plain fact approach to legal positivism: according to Dyzenhaus, a judge who accepts the positivist analysis of law must be a plain fact judge. Now, Dworkin's plain fact view of law may not hold up a mirror to all legal positivists. But it certainly does to some. The two claims which Dworkin attributes to the plain fact view more or less capture two central ideas in Joseph Raz's analysis of law. One is Raz's so-called "sources thesis". It holds that the existence and content of every one of a community's legal rules (norms or standards), of all its law in other words, can be determined by reference only to social facts without any reliance on moral considerations.[17] The other is Raz's insistence that the law so identified will on occasion be indeterminate. It will, at times, fail to identify a single solution as the right solution to a legal question.[18]

Will any legal positivists recognise themselves in Dyzenhaus's plain fact approach? Surely not. Most positivists would reject at least two of its four tenets. In fact, the plain fact approach described by Dyzenhaus seems to have less in common with legal positivism than it does with one of legal positivism's most powerful rivals, namely Dworkin's "law as integrity". Dyzenhaus's plain fact approach is really just a corrupted version of Dworkin's analysis of law. Or to put it another way, Dyzenhaus's plain fact judge is really just a flawed version of Dworkin's ideal judge, "Hercules".[19]

It is impossible here to provide a proper exegesis of Dworkin's jurisprudence. For present purposes, however, it is sufficient to highlight four theses which seem to be at its core. The first is Dworkin's so-called "right-answer thesis", which holds that a community's law determines one decision as correct in every legal dispute.[20] The second thesis deals with the authority of a community's law. According to Dworkin, a community's law provides judges with reasons for decision which are only exceptionally defeasible by contrary moral reasons.[21] The third thesis concerns the identification of a community's law and the fourth has to do with adjudication. Neither of these theses can be understood, however, unless we first come to terms with Dworkin's notion of "constructive interpretation".

Constructive interpretation, in the legal context, comprises the following three stages.[22] At the "pre-interpretive" stage one establishes what I will call the community's "legal record". This is to be done by giving the law-making acts of the community's law-makers (typically statutes enacted by the legislature and decisions handed down by the courts) the meaning the law-makers intended

[17] Joseph Raz, *The Authority of Law* (Oxford: Clarendon Press, 1979) at pp. 37–52, 53; "The Problem about the Nature of Law" (1983) 21 *Univ of Western Ontario LR* 203 at 214, 217–18; "Authority, Law and Morality" (1985) 68 *The Monist* 295 at 295–6.

[18] Raz, *The Authority of Law*, n. 17 above, at pp. 70–7.

[19] See Dworkin, *Taking Rights Seriously*, n. 1 above, at p. 105; *Law's Empire*, n. 8 above, at p. 239.

[20] Dworkin, "Hard Cases" and "Can Rights be Controversial?" in *Taking Rights Seriously*, n. 1 above; "Is There Really No Right Answer in Hard Cases?" in *A Matter of Principle* (Oxford: Clarendon Press, 1985); *Law's Empire*, n. 8 above, at pp. 266–75.

[21] Dworkin, *Law's Empire*, n. 8 above, at pp. 108–13, 218–19.

[22] See Dworkin, *Law's Empire*, n. 8 above, at pp. 47–8, 52–3, 65–8, 225–58.

them to have (normally their ordinary meaning – their meaning as fixed by the relevant linguistic conventions).[23] Then, at the "interpretive" stage, one establishes the scheme of principles which satisfies the following two conditions. One: it adequately "fits" the legal record, i.e., someone holding the scheme of principles could have been led thereby to enact most of the legal record, including its most important parts. Two: the scheme of principles provides a better "justification" of the legal record than does any other scheme which also adequately fits the legal record. How good a justification of the record a scheme of principles provides depends, among other things, on its proximity to principles of justice (morality) and the closeness of its fit with the record. Finally, at the "post-interpretive" stage, one folds the scheme of principles back on the legal record which provides the principles' foundation, in order to draw a distinction between those parts of the legal record which are sound and those which are "mistaken". As explained above, the scheme of principles which is identified at the interpretive stage need not fit all of the legal record, but only "the bulk" thereof. Dworkin accepts that, normally, there will be some parts of the legal record which the scheme of principles will not fit. These, says Dworkin, are to be treated as mistakes.

What does this have to do with either the identification of law or adjudication? Well, according to Dworkin, the scheme of principles which adequately fits and best justifies a community's legal record is part of that community's law.[24] And this scheme of principles is to guide judges in deciding cases.[25] As regards the legal record, Dworkin's remarks are at best ambiguous, at worst contradictory.[26] Sometimes he seems to say that those parts of the legal record which have been identified as mistakes fall outside of a community's law and should not guide adjudication. This entails, of course, that every judicial decision (including ones simply applying the legal record) must involve constructive interpretation. On other occasions, however, Dworkin seems to suggest that all of a community's legal record, whether mistaken or not, is part of the community's law. And he seems to say that, in so far as the legal record is determinate, adjudication is to proceed without adverting to the principles which fit and justify the legal record. The court should simply make the decision required by the record. Only when the legal record provides no solution to a case is the court to seek guidance from these principles. Constructive interpretation, in other words, should only kick in at the point where the legal record runs out.

Enough has been said about Dworkin's jurisprudential views to enable an assessment of the assertion that, on close inspection, Dyzenhaus's plain fact judge reveals himself to be a half-baked Hercules. Two ideas are clearly shared

[23] See Dworkin, *Law's Empire* at pp. 227, 338; "Bork's Jurisprudence" (1990) 57 *Univ of Chicago LR* 657 at 661–2; *Life's Dominion* (London: Harper Collins, 1993) at pp. 133–6; *Freedom's Law* (Oxford: Oxford University Press, 1996) at 8–10; "The Arduous Virtue of Fidelity: Originalism, Scalia, Tribe, and Nerve" (1997) 65 *Fordham LR* 1249 at 1251–5.

[24] See Dworkin, *Law's Empire*, n. 8 above, at pp. 225, 227.

[25] See ibid., pp. 112, 218–19.

[26] See Dworkin, *Taking Rights Seriously*, n. 1 above, at pp. 118–23.

by both Dworkin and Dyzenhaus's plain fact judge. One is that the law of a community provides solutions to all disputes. The other is that the law of a community has legitimate authority (though, as we have seen, Dworkin may accord law marginally less authority than does Dyzenhaus's plain fact judge). In my view, Dyzenhaus's plain fact judge also shares a third idea with Dworkin. Namely, although we identify some of a community's law by giving its law-making acts their intended meaning, we do not so identify all of it.

This requires some explanation. According to Dyzenhaus, a plain fact judge identifies his community's law by employing a two-fold intentional test. The judge starts by attributing to every law-making act the meaning the law-maker intended it to have.[27] If, however, the meaning so attributed is uncertain or unclear, the judge is to have recourse to a second set of intentions. The content of these intentions is to be established by a particular form of interpretative reasoning. Namely, the judge is to identify the "moral ideas" or "ideology" or "overall design" that most probably motivated the law-makers to perform their law-making acts.[28] There is, I believe, another way of describing the latter interpretative process. To wit, the judge is to identify the scheme of principles which "best fits" the community's legal record (in the sense that a person acting on that scheme of principles is more likely to have enacted the legal record than a person acting on any other scheme of principles). If this is a permissible redescription, then Dyzenhaus's plain fact judge is really just a Dworkinian Hercules *sans* "best justification". That is, if you take Dworkin's analysis of law and remove just one ingredient, namely "best justification" as a condition for the identification of a community's legal principles, then, bingo, you have Dyzenhaus's plain fact judge.

If my understanding of Dyzenhaus's plain fact approach is correct, then Dyzenhaus's project is a particularly ambitious one. Dyzenhaus's critique of positivism depends on there being a connection between positivism and the plain fact approach. It depends on the claim that judicial acceptance of the former must lead to judicial acceptance also of the latter. As we have seen, however, whilst the plain fact approach described by Dyzenhaus contains little that would be amenable to the positivist, it contains much that would be congenial to Ronald Dworkin, legal positivism's most fierce and enduring critic. In order to succeed with his critique, Dyzenhaus thus faces the difficult task of showing that legal positivism does not know itself. He must show that, notwithstanding legal positivists' protests to the contrary, their analysis of law really does inevitably lead judges who accept it also to accept the plain fact approach.

Dyzenhaus is fully aware of the difficulty he faces. But he is confident that he has the argument to overcome it. In outline, the argument is as follows. One: the central tenet of contemporary positivism is "that law is law which meets a sources test",[29] that is, all of a community's law can be identified by reference to

[27] Dyzenhaus, *Hard Cases in Wicked Legal Systems*, n. 4 above, at pp. 56–8.
[28] Ibid., p. 57.
[29] Ibid., pp. 213, 239.

social facts only, without any resort to moral considerations. Two: this means that a judge who endorses legal positivism must also endorse the positivist rationale for the sources test.[30] Three: if we take a closer look at the positivist rationale for the sources test, we discover that a judge can endorse this rationale only if the judge also accepts the following three claims.[31] The law has legitimate authority. The law is to be identified by reference not only to the meaning the law-makers intended their acts to have but also to the moral ideas or ideology which moved the law-makers to their acts. Judges have no discretion. Four: to accept these three claims together with the sources test is to adopt the plain fact approach. Hence: a positivist judge is a plain fact one.

The above argument stands or falls by its third point. Is the point valid? Does acceptance of the positivist rationale for the sources test really commit one also to the other three claims constitutive of the plain fact approach? Dyzenhaus believes so. I believe not. In order to defend my take on this against Dyzenhaus's contrary one I attempt, in the next section, to set out what I take to be the best account of the positivist rationale for the sources test. In the section thereafter I explain why, so understood, the positivist rationale does not have the consequences that Dyzenhaus claims for it.

II. THE ARGUMENT FROM MORAL INTELLIGIBILITY

Before diving into an account of the positivist rationale for the sources test, I would like to make two prefatory remarks. The first is to caution against too easy an association of legal positivism and the sources test. The sources test, or sources thesis (as I will henceforth refer to it), is the progeny of Joseph Raz. Raz is, without a doubt, one of the foremost expositors of contemporary legal positivism. But not all who fly the flag of contemporary legal positivism have declared their allegiance to Raz's sources thesis. Jules Coleman, Philip Soper, David Lyons and, more recently, Wil Waluchow have espoused an account of law that is clearly incompatible with the sources thesis.[32] Yet they insist that their account, which has come to be known as "incorporationism", "soft positivism" or "inclusive positivism", remains true to the central tenets of legal positivism.

This suggests that Dyzenhaus's critique of legal positivism can at best hope to achieve a partial success. It is essential to Dyzenhaus's critique that positivism connect with the plain fact approach. According to Dyzenhaus, as we have seen, it is the positivist rationale for the sources thesis which bridges the gap between

[30] Dyzenhaus, *Hard Cases in Wicked Legal Systems* at p. 239.

[31] Ibid., p. 239.

[32] Jules Coleman, "Negative and Positive Positivism", Philip Soper, "Legal Theory and the Obligation of a Judge: The Hart/Dworkin Dispute", both in Marshall Cohen (ed.), *Ronald Dworkin and Contemporary Jurisprudence* (London: Duckworth, 1983); David Lyons, "Principles, Positivism, and Legal Theory" (1977) 87 *Yale LJ* 415 at 423–4; W.J. Waluchow, *Inclusive Legal Positivism* (Oxford: Clarendon Press, 1994).

legal positivism and the plain fact approach. Since the soft positivists reject the sources thesis and since Dyzenhaus provides no alternative means of linking soft positivism to the plain fact approach, it seems that soft positivism must fall outside the ambit of Dyzenhaus's critique.

Dyzenhaus is by no means ignorant of soft positivism. But he feels able to discount it, it would seem, on the grounds that it is merely an "offshoot", falling outside the mainstream, of contemporary legal positivism.[33] His reason for so regarding it, I suspect, stems in part from his assumption that the sources thesis has the support of H.L.A. Hart, indisputable doyen of modern positivism.[34] It is understandable that Dyzenhaus should have assumed this, given the fact that his understanding of Hart's jurisprudence was based on the first edition of *The Concept of Law*. Indeed, it would appear that Ronald Dworkin read the first edition in the same way. For the "pedigree test" which Dworkin attributed to Hart bears a striking resemblance to Raz's sources thesis.[35] Since the publication of Hart's "Postscript" to the second edition of *The Concept of Law*, however, this understanding of Hart's views is no longer sustainable. For, in the postscript, Hart explicitly aligns himself with the soft positivists.

So it would be wrong to think of Raz's sources thesis as constituting the centre, and of the soft positivism of Coleman, Lyons, Soper and Waluchow as being at the margins, of modern legal positivism. If anything, the opposite may be the case. It follows that, if Dyzenhaus wishes to provide a comprehensive refutation of contemporary legal positivism, he will have to supplement his present critique. He must either provide an independent argument against soft positivism or explain just why soft positivism does not warrant its appellation.

The second remark to some extent flows from the first. According to Dyzenhaus, it is the positivist rationale for the sources thesis which forges the link between positivism and the plain fact approach. In order to test this claim I intend, in this section, to determine just what that rationale is. As we saw a moment ago, however, whilst the sources thesis is fiercely advocated by Joseph Raz, it has by no means earned the acceptance of all his fellow positivists. This suggests that, if we are to discover a rationale for the sources thesis, we should look first of all to Raz's writings. Indeed, a search through Raz's many publications yields not one, but several, arguments for the sources thesis.[36] I will not attempt to paraphrase them here. Instead, I will present what I take to be the best possible argument for the sources thesis. I will call it the "argument from moral intelligibility". The argument is not Raz's, as it is not identical with

[33] Dyzenhaus, *Hard Cases in Wicked Legal Systems*, n. 4 above, at p. 24.

[34] Ibid., p. 24.

[35] Dworkin, *Taking Rights Seriously*, n. 1 above, at p. 17.

[36] See Raz, *The Authority of Law*, n. 17 above, at pp. 41–5, 48–52; *The Concept of a Legal System*, 2nd edn. (Oxford: Clarendon Press, 1980) at p. 216; "Authority, Law and Morality", n. 17 above; "The Problem about the Nature of Law", n. 17 above, at 211–18; "The Relevance of Coherence" (1992) 72 *Boston Univ LR* 273 at 292–7; "Intention in Interpretation" in Robert George, (ed.) *The Autonomy of Law* (Oxford: Clarendon Press, 1996) at pp. 256–62; "On the Nature of Law" (1996) 82 *Archiv fur Rechts- und Sozialphilosophie* 1 at 17–18.

anything he offers. But it is most certainly Razian. For all its key ideas are culled from his work.

Point one in the argument from moral intelligibility is unlikely to elicit many objections. It holds that, since law is a social practice constituted in part by the beliefs (attitudes, values and so on) of its participants, a proper understanding of law must be belief-centred. It must be sensitive to the beliefs that those who participate in the practice of law have regarding it. This idea can be traced back to Hart's emphasis on "the internal point of view" in his *The Concept of Law*.[37] Today it enjoys almost universal acceptance among legal philosophers, whether they be proponents or detractors of legal positivism.[38] So I will not dwell on it.

Point two in the argument from moral intelligibility identifies two beliefs that an account of law cannot possibly ignore. One is the belief that it matters who a community's law-makers are and whether they apply their minds to their law-making. The other belief is rather more complex. Namely: it is morally justified for a person to perform an action required by the law of his community even when, but for the law, the person would have regarded the action as morally undesirable. This belief may be broken down into three parts. First, the law provides people with new reasons for action. Secondly, the reasons which the law provides by and large prevail over countervailing moral reasons. Thirdly, the reasons which the law provides are moral reasons. In what follows I will refer to this complex belief simply as the belief that law is morally binding, or as the belief that law has legitimate authority.

Some people may take issue with my imputation to law's participants of the belief that law is morally binding. Their unhappiness is unlikely to relate to the first and second constituents of this belief. But what about the third? Do law's participants truly believe that it is morally justified for people to prioritise law in this way? Hart believed not. Law's participants are not, in his view, "committed to a *moral* judgment that it is morally right to do what the law requires".[39] On this issue, however, Hart's is a lone voice. And rightly so. As Raz explains, it is a necessary truth that most people most of the time act, and desire others to act, in ways they believe to be good or valuable. If, therefore, law's participants accord law a preeminence over contrary moral reasons, it must be that, by and large, they believe this to be morally justified.[40]

According to point three in the argument from moral intelligibility, for an understanding of law to be sensitive to the beliefs of participants about the law,

[37] Hart, *The Concept of Law*, n. 1 above, at 55–7, 88–91, 98–9, 242.

[38] See for example Dworkin, *Law's Empire*, n. 8 above, at pp. 13–14; John Finnis, *Natural Law and Natural Rights* (Oxford: Clarendon Press, 1980) at p. 3; Neil MacCormick, "Law, Morality and Positivism" in Neil MacCormick and Ota Weinberger, *An Institutional Theory of Law* (D. Reidel, 1986) at pp. 130–1; Raz, "The Relevance of Coherence", n. 36 above, at 292–3, 295, "Authority, Law and Morality", n. 17 above, at 321–2, "Intention in Interpretation", n. 36 above, at 262.

[39] Hart, *The Concept of Law*, n. 1 above, at p. 203.

[40] Raz, "Intention in Interpretation", n. 36 above, at 261; "On the Autonomy of Legal Reasoning" (1993) 6 *Ratio Juris* 1 at 8–9. For a view in substantial agreement with this, see Neil MacCormick, *Legal Reasoning and Legal Theory* (Oxford: Clarendon Press, 1978) at pp. 63–4, 139–40.

it need not render those beliefs true but must render them intelligible. We have some idea of what it is for a belief to be true. But what exactly is it for a belief to be intelligible?

Raz proffers this explanation.[41] The belief that an object has a particular property can be false in two ways. It could be that the object by its very nature is incapable of possessing that property. If so, then to believe that the object has the property would be to make a conceptual mistake. It would be to misunderstand the nature of the object or the property or both. For example, a tree cannot possibly be a promisee. To believe that one owes a tree an obligation by virtue of a promise one has made it would thus involve a conceptual confusion about trees and/or promises. Alternatively, it could be that, though nothing in the nature of the object and the property precludes the former from possessing the latter, it in fact does not. In this event, the belief that the object has the property involves no conceptual error, but is simply wrong on the facts. For example, it is quite possible that John be a promisee. But maybe John has never received any promises. If so, the belief that John is a promisee is false, but it need not involve any conceptual confusion.

What does this have to do with the notion of intelligibility? Well, according to Raz (as I understand him), for a belief to be intelligible it is necessary that it be conceptually sound. The belief that an object has a particular property is intelligible, therefore, only if one makes no conceptual mistake in holding it. Only if, that is, the object is not by its very nature precluded from possessing the property. At the same time, it is not a necessary condition for a belief's intelligibility that it be correct on the facts. An intelligible belief may be a false one.

The fourth point in the argument from moral intelligibility claims that the two beliefs identified in point two are intelligible only if law-makers have the ability to make law that reflects their moral judgement. Consider, in the first place, the belief that it matters who a community's law-makers are and whether they apply their minds to their law-making. The main reason that people care about the make-up of their community's law-making bodies is that they are concerned about the law emanating from those bodies. Of course, a concern about the law provides a reason to care about the law-makers only if who the law-makers are makes a difference to what law one gets. More than that, it must be that the law emerging from the law-making bodies by and large reflects the views, prejudices and moral judgements of those who comprise the law-making bodies. Clearly, therefore, if popular concern about the composition of law-making bodies is to make any sense—if such concern is to be intelligible—it must be that these bodies are at the very least capable of making law in accordance with their views, prejudices and judgements. A similar inference can be drawn from the fact that people expect their law-makers to make law only after proper deliberation. If law-makers are incapable of making law that reflects their moral judgement, why should anyone care whether law-makers act on

[41] Raz, "Authority, Law and Morality", n. 17 above, at 300–2.

good or bad judgements, conscientiously or flippantly, with or without prior deliberation?[42]

Consider, in the second place, the belief that law is morally binding or has legitimate authority. This belief, to recall, holds that it is morally justified for people to perform actions required by the law even when, in the absence of the law, they would have judged those actions to be morally unacceptable. We can only establish what is required for this belief to be intelligible, if we first determine what is required for it to be true. Some writers, most notably Robert Paul Wolff, have argued that this belief is necessarily false. Why so? Because morality requires one always to act autonomously. To act autonomously is to act upon one's own moral judgement. To follow the law rather than the moral convictions one holds independently from the law is, however, to abandon or surrender one's own moral judgement.[43]

Wolff's argument has been persuasively refuted by Raz. Essential to Raz's refutation is a distinction he draws between "conforming" and "complying" with a reason for action.[44] If p is a reason for me to perform some act, then I conform with reason p if I perform the act in the appropriate circumstance. I comply with reason p if I perform the act in the appropriate circumstance, and I do so *for the reason that p*, rather than for any other or no reason. According to Raz, conformity with moral reasons matters more than compliance with them: "the important thing is that the act for which the reason is a reason gets done".[45] Where p is a moral reason for me to perform some act, the important thing is that I perform the act, not that I perform it for the reason that p. Since conformity with moral reasons is important, "[o]ne has reason to do whatever will facilitate conformity with [moral] reason".[46] This explains why, under normal circumstances, one should try to comply with moral reasons. For, if one complies with a reason, one necessarily conforms with it. Most of the time, our attempts at compliance are successful. So a good way to achieve conformity with moral reasons is by trying to comply with them. The attempt at compliance, on this account, has mainly instrumental value. As Raz writes: "Normally compliance matters only because attempting to comply is the only reliable route to conformity".[47] Sometimes, however, conformity with moral reasons is better served by not trying to comply with those reasons directly. The attempt to do so may make conformity less likely. Conformity with moral reasons may be better achieved by compliance with some other set of reasons. Since it is conformity

[42] See Raz, "The Relevance of Coherence", n. 36 above, at 289, 295; "Intention in Interpretation", n. 36 above, at 258–9.

[43] See Robert Paul Wolff, "The Conflict between Authority and Autonomy" in Joseph Raz (ed.), *Authority* (Oxford: Basil Blackwell, 1990).

[44] Joseph Raz, *Practical Reason and Norms*, 2nd edn. (Princeton: Princeton University Press, 1990) at p. 178.

[45] Ibid., p. 180.

[46] Ibid., p. 182.

[47] Ibid., p. 190.

that matters, one should in this event abandon the attempt at direct compliance and try instead to comply with the alternative set of reasons.

Raz's distinction between conformity and compliance, and his rejection of the idea that trying to comply with moral reasons is more important than conforming with them, show how it can be morally justified for people to privilege the law of their community over their contrary moral convictions. What matters, ultimately, is that people's actions conform with the moral reasons which apply to them. In the final analysis, we want people's actions to be morally sound. It may be, however, that this is best achieved by an indirect route. Perhaps the members of a community are more likely to perform actions that conform (or likely to perform actions that better conform) with the moral reasons which apply to them if they do not try to comply with those reasons directly, but instead abide by the requirements of their community's law. If this is indeed the case, then it is morally justified for the community's members to act as the law requires even when, but for the law, their moral judgement would have counselled against so acting.[48]

The upshot of this is as follows. The popular belief that law is morally binding or has legitimate authority is true if and only if people do better by following the law than by following their independent moral judgement. How could this condition possibly be satisfied? What, that is, could possibly bring it about that a community's law provides a more reliable guide to the moral reasons applicable to people's actions than do people's own assessments of those reasons? As Raz points out, there is only one plausible answer to this question. Namely, the community's law-makers have a moral judgement superior to that of the community members, and the law reflects that superior judgement.[49]

Of course, our concern is with the intelligibility rather than the truth of the belief that law is morally binding. What is required for this belief to be intelligible? Surely, as a bare minimum, the law-makers must have the ability to make law in accordance with their moral judgement. Law-makers must, that is, be able to express in the law their view on how their subjects should act. For, as was explained in point three, a belief is intelligible if one makes no conceptual mistake in holding it. More specifically, the belief that an object has some property is intelligible only if the object is at the very least capable of possessing that property. The belief that law is morally binding is thus intelligible only if law is capable of being morally binding. It must be possible, in other words, that the law improves on the moral judgement of its subjects. But how could this be possible, unless the law-makers are able to make law in conformity with their moral judgement?

[48] This is simply a reformulation of Raz's "normal justification thesis". See Raz, "Authority, Law and Morality", n. 17 above, at 299; *The Morality of Freedom* (Oxford: Clarendon Press, 1986) at p. 53.

[49] See Raz, "Authority, Law and Morality", n. 17 above, at 303–5, 320–1, "Intention in Interpretation", n. 36 above, at 258–9, 282 n.10.

According to point five in the argument from moral intelligibility, a community's law-makers will be able to make law that reflects their moral judgement if and only if the following two conditions are satisfied. One: for any law, there exists some act such that, if the law-makers perform it, that law (and only that law) will be created. Two: the law-makers are able to determine, in advance, which acts create which laws. Satisfaction of these two conditions ensures that a community's law-makers have control over the law they make. If they decide that there should be a law that P, they can go about making just that law. All they have to do is find out what act they need to perform in order to create this law, and then perform it. Likewise, if the law-makers do not want to make a law that P, they have the means to avoid doing so. They must just ensure that they do not perform any acts, performance of which would create the law that P.

The argument from moral intelligibility is completed by its sixth point. It goes as follows. If the two conditions set out above are to be satisfied, if the law-makers are to have control over the law they make, the law of a community must be identified by reference to social fact and not by recourse to morality. If point six is correct, the sources thesis is vindicated. Points one through to five of the argument from moral intelligibility establish that a theory of law is unacceptable if it has the result that a community's law-makers lose control over the law they make. For, as point five shows, if the law-makers have no control over the law they make, they will also lack the ability to express their views in the law. As point four makes clear, if the law-makers lack this ability, the belief that law is morally binding (and the belief that the composition and decision-making of the law-makers matter) will be morally unintelligible. Points one, two and three show that an understanding of law which renders these beliefs unintelligible will necessarily be invalid. For it will be insensitive to the beliefs held by law's participants and sensitivity is a requirement for an understanding of law's acceptability. Now, point six in the argument from moral intelligibility alleges that, if a theory of law involves moral considerations in the identification of law, then it denies control over the law to law-makers. The combination of this with points one to five yields the conclusion that a theory of law is unacceptable if its test for law involves moral considerations. This is the sources thesis.

Of course, it still has to be shown that point six is correct. It is easy enough to see why the involvement of morality in the determination of a community's law frustrates the second of the two conditions set out in point five. Even if we accept (as I do) that there are knowable moral truths, no law-maker has perfect moral knowledge. It follows that, to the extent that morality determines the law created by the law-makers' law-making acts, the law-makers will not know just what law they are creating. The law-makers will be unable to achieve certainty about the law-making consequences of their law-making acts. Note that I am not here relying on the controversial nature of moral claims. This will make it difficult for law-makers to predict their subjects' beliefs about their community's law. But it has no impact on the law-makers' ability to ascertain the law created by their acts of law-making.

What about the first condition laid down in point five? Well, certainly, if conformity with morality is a necessary feature of a community's law, then the first condition will be frustrated. Certain laws, namely those which conflict with morality, will be beyond the law-making competence of the law-makers. Where the law-makers judge a law to be desirable, but the law does not conform with morality, no act will be available to the law-makers for the creation of that law.

The foregoing shows that a theory of law denies law-makers control over the law they make if it claims or entails that morality necessarily has a part to play in the determination of a community's law. This goes a fair distance towards establishing that point six in the argument from moral intelligibility is sound. But it does not get us quite far enough. It is one thing for a theory of law to assert that morality necessarily has a part in the identification of a community's law. It is quite another for a theory to claim for morality a merely contingent role in the identification of law, to claim no more than that morality may or can play a part in law's identification. For point six to be established, it has to be demonstrated that a loss of control on the part of law-makers results from any involvement of morality in the identification of law, whether that involvement be necessary or contingent. I have shown the former. But I have yet to show the latter.

The claim that morality may or can (rather than necessarily does) figure in the identification of a community's law is a central tenet of soft positivism, as that is espoused by Jules Coleman, David Lyons, Philip Soper and, more recently, by W.J. Waluchow and H.L.A. Hart.[50] According to soft positivism, whether morality does or does not have a role in the determination of a community's law depends on social facts. For, in order to establish a community's law, one must necessarily start with its "source-based" law. That is, one must start with law that can be identified by reference only to social facts, such as the practices of law-applying officials, the law-making acts of law-making institutions and the conventions of meaning pertinent to those law-making acts. It could be that this source-based law is exhaustive of a community's law. If so, the community's law can be identified without resort to moral considerations.

It is equally possible (and more usual), however, that the source-based law makes use of (or invokes) moral terms and concepts, such as fairness, justice and equality. So, for example, the South African Bill of Rights (the text enacted by the South African Constitutional Assembly, given its conventional meaning) provides for rights to equality, dignity, privacy and freedom of expression, but allows for their infringement so long as the infringement is reasonable and justifiable in an open and democratic society based on human dignity, equality and freedom. In this event, say the soft positivists, morality is incorporated into the community's law. The community's law consists not only of its source-based

[50] Coleman, "Negative and Positive Positivism", n. 32 above; Soper, "Legal Theory and the Obligation of a Judge: The Hart/Dworkin Dispute", n. 32 above; Lyons, "Principles, Positivism, and Legal Theory", n. 32 above, at 423–4; Waluchow, *Inclusive Legal Positivism*, n. 32 above; Hart, *The Concept of Law*, n. 1 above, in the "Postscript".

law, but also of the correct account (the best theory) of the moral concepts figuring in the source-based law. This can work at any level in the hierarchy of law. If a community's rules of recognition (be they "ultimate" practice-based ones or "inferior" ones identified by the ultimate ones) include moral criteria like fairness and justice, then the best accounts of fairness and justice are incorporated into the rules of recognition and hence into the community's law.[51] Equally, if a municipal regulation requires pubs to keep reasonable hours, then the best account of reasonableness is incorporated into the regulation and so (again) into the law.

Soft positivism presents the following difficulty for my attempt to justify the sources thesis. In order for the sources thesis to be justified by the argument from moral intelligibility, the argument's sixth point must be sound. That is, it must be true that law-makers are denied control by a theory of law which involves moral criteria in the identification of law. As we have seen, if a theory of law makes morality a necessary determinant of a community's law, it entails such a loss of control. But soft positivism accords morality a merely contingent role in the identification of law. More than that, the role morality has in law's identification is contingent upon source-based law. Morality has a part to play in a community's law only if the community's source-based law incorporates it.

Why does this threaten point six in the argument from moral intelligibility? Well, because source-based law clearly is under the control of a community's law-makers (since this idea is essential to the justification of the sources thesis, its proponents cannot possibly deny it). But, if source-based law is under the control of a community's law-makers and the incorporation of morality into the community's law is contingent upon its source-based law, then it surely follows that the community's law-makers retain control over the community's law. If this is right, if soft positivism does not after all entail that law-makers lose control over their community's law, then point six in the argument from moral intelligibility is incorrect. It is not so that law-makers are denied control over the law whenever morality figures in the test for identifying law.

If point six in the argument from moral intelligibility really is incorrect, if a theory of law really may involve morality in its test for law without compromising law-makers' control over the law, then the argument from moral intelligibility cannot vindicate the sources thesis. But does soft positivism really preserve law-makers' control over the law? I believe not. Assume that the playing of stereos at full volume in a park constitutes a disturbance to others' enjoyment of the park and that such a disturbance constitutes unreasonable behaviour in the park. Does it follow that the following three rules are identical? One: no unreasonable behaviour in the park. Two: no disturbance of others' enjoyment of the park. Three: no one is to play a stereo at full volume in

[51] This means that, in such a community, the answer to the question "What is the law of the community?" is not unsettled, as it would be if the sources thesis were valid, though it may be controversial. The community's law is simply that which is in fact fair and just according to the correct understanding of these concepts.

the park. Clearly not. Nor would any serious legal theorist contend it. Now, as was made clear in point five above, a community's law-makers cannot be said to have control over the law of their community unless, for any law, there exists some act such that, if the law-makers perform it, that law (and only that law) will be created. This means that, for a law-maker to have control over the law of his community, he must be able to make any one of the above three rules law, without the other two becoming law at the same time. If soft positivism is correct, however, then a law-maker cannot make the first rule law without also making the other two rules law. Nor can a law-maker make the second rule law without the third.

What this shows is that, contrary to first appearances, soft positivism does in fact place a severe constraint on the control that law-makers have over their community's law. Soft positivism makes it impossible for law-makers to limit the law they make to what Ronald Dworkin calls "abstract" rules (or principles). As an example of an abstract principle, Dworkin gives the principle that everyone should be treated equally.[52] If soft positivism is sound, there is no way that a community's law-makers could enact just that principle. Of course, they could make only that principle part of their community's source-based law. This would be achieved by promulgating a law-making text the ordinary meaning of which is simply that everyone should be treated equally. But, if soft positivism is sound, then the law includes also the correct moral understanding of this principle. It could be that, correctly understood, this principle entails the more "concrete" principle that those who were treated unequally in the past because of their membership of certain groups should now receive preferential treatment in certain contexts, and that the latter in turns entails the even more concrete principle that, in South Africa, women and blacks should be preferentially employed over white men in the public sector. According to the soft positivists, the incorporation of these concrete principles is an inevitable consequence of the law-makers' enactment of the more abstract one. This being so, there is no way for the law-makers to limit their law-making to the abstract principle.

III. INTENTION, AUTHORITY AND DISCRETION

David Dyzenhaus believes that a judge who endorses the sources thesis for the identification of a community's law must also endorse the plain fact approach. For a commitment to the sources thesis entails a commitment to its justification or rationale. Acceptance of the rationale for the sources thesis in turn compels acceptance of three further claims: that a community's law is to be identified so as to reflect the law-makers' intentions (including the moral ideology which motivated the law-makers), that the law so identified has legitimate authority

[52] Dworkin, "The Forum of Principle" (1981) 56 *New York Univ LR* 469 at 489; "Bork's Jurisprudence", n. 23 above, at 665.

(in other words, if the law provides a solution to a case, the courts are morally bound to follow it) and that judges have no discretion (that is, judges never decide cases on non-legal grounds). Together with the sources thesis, these three claims are constitutive of the plain fact approach.

In the previous section I set out, in some detail, a rationale for the sources thesis. I called it the argument from moral intelligibility. In my view, the argument from moral intelligibility validates the sources thesis. If this is right, then Dyzenhaus's attempt to connect the sources thesis to the plain fact approach will succeed only if the argument from moral intelligibility entails the three claims outlined above. My aim in this section is to show that it does not. I should state at the outset that I will only be presenting argument against the first two claims. For it seems to me that the invalidity of the first two claims entails the invalidity of the third as well.

Let us start with the idea that, if a judge identifies his community's law in accordance with the sources thesis, and does so on the basis of the argument from moral intelligibility, he must also view the law of his community as having legitimate authority. Clearly, this is not true of persons generally. A legal theorist observing a community's legal system from the "outside", as it were, can at the same time acknowledge that the community's law is to be identified in accordance with the sources thesis, for the reasons provided by the argument from moral intelligibility, and deny that the law of the community is morally binding. There is nothing contradictory in this. One premise of the argument from moral intelligibility is that law's participants believe it to possess legitimate authority. Another premise is that this belief is intelligible: law's participants are not conceptually confused in holding this belief. It is not, however, a premise of the argument from moral intelligibility (nor entailed thereby) that this belief is true.

It may be felt that judges are a special case, however. In a certain sense this is correct. Judges are legal participants *par excellence*. Hence, were law's authority denied by a sufficient number of judges, the sources thesis would probably be invalid. Or at least, its validity could not be based on the argument from moral intelligibility. For the argument's empirical foundation, namely the belief by law's participants that law has legitimate authority, would have been eroded.

Quite obviously, however, the fact that one endorses an argument founded on a particular empirical premise does not in any way oblige one to maintain the conditions which make that premise true. A judge's acceptance of the argument from moral intelligibility does not therefore impose on the judge any responsibility to maintain the belief that provides the argument's empirical premise, namely that law has legitimate authority. Alternatively put: although a positivist judge must believe that the conditions by virtue of which positivism is valid obtain, he is under no compulsion to sustain those conditions.

So the fact that a judge endorses the sources thesis, and does so on the grounds of the argument from moral intelligibility, does not preclude the judge from denying that his community's law lacks legitimate authority. This does not mean that there may not be other reasons for judges to treat their community's

law as authoritative. Judges generally undertake to the public that they will apply their community's law. The public generally expects judges to do just that. These are good reasons for judges to treat their community's law as morally binding. But they are wholly independent from the argument from moral intelligibility. This means that a judge who, on the one hand, endorses the sources thesis and, on the other, rejects law's authority may well be acting improperly. But he cannot be accused of acting inconsistently with his commitment to the sources thesis.

Let us move on to the idea that, if a judge endorses the sources thesis, and does so on the basis of the argument from moral intelligibility, he must also embrace intentionalism. That is, the judge has no choice but to take the view that a community's law is to be identified so as to reflect the intentions of the community's law-makers in performing their law-making acts. Of course, when law-makers perform a law-making act, they typically have two intentions rather than one. They have the intention to make a particular legal rule. And they have the intention to achieve a particular result by making the rule. To differentiate these intentions I will call the former the law-makers' law-making intention and the latter their further intention.[53] This is similar to the distinction between what a speaker intends to say by an utterance and what the speaker intends to bring about by saying it. For example, when I utter "My, but it's cold in here", I (normally) intend to say just that it is cold in here. But I may also intend my saying that to get you to close the door you have just left open.[54]

Corresponding to the distinction between further and law-making intentions, there is also a distinction that can be drawn between two kinds of intentionalism. There is intentionalism in the full-blooded sense in which it is usually understood. This advocates a role for law-makers' further intentions in the identification of a community's law. It is possible, however, to conceive of intentionalism in a more anaemic form which pays no attention to law-makers' further intentions. Intentionalism so understood claims no more than that a community's law is to be identified so as to reflect the law-making intentions of the community's law-makers.

The plain fact judge, as Dyzenhaus describes him, clearly is an intentionalist.[55] At a glance, it would appear that Dyzenhaus's plain fact judge is also a full-blooded intentionalist. For he identifies his community's law by the following two-stage process. He starts by identifying the law in accordance with the law-makers' law-making intentions. If, however, the law so identified turns out to be indeterminate, he supplements it with law identified in accordance with the law-makers' further intentions. Yet, when we take a close look at the way in which

[53] On further intentions, see Andrei Marmor, *Interpretation and Legal Theory* (Oxford: Clarendon Press, 1992) at pp. 165–72; Raz, "Intention in Interpretation", n. 36 above, at 274–7.

[54] On the distinction between what speakers intend to say by their utterances and what they intend to achieve as a result of their having said it, see Dworkin, *Life's Dominion*, n. 23 above, at pp. 133–5.

[55] Dyzenhaus, *Hard Cases in Wicked Legal Systems*, n. 4 above, at pp. 56–8.

Dyzenhaus's plain fact judge establishes what the further intentions of the law-makers were, we become aware of the following difficulty. As we saw in an earlier section, Dyzenhaus's plain fact judge identifies law-makers' further intentions by means of an interpretative process strongly reminiscent of Ronald Dworkin's test of "fit". Dyzenhaus's plain fact judge "attributes" further intentions to law-makers, on the grounds that they provide "the best historical explanation" for the law-makers' record of law-making.[56] Now, one way of understanding this is that "the best historical explanation" is constitutive of the law-makers' further intentions. If this understanding is correct, then Dyzenhaus's plain fact judge is not a full-blooded intentionalist after all. For the full-blooded intentionalist has no interest in coherentist abstractions. His concern is only with aims and goals actually entertained by the law-makers: real intentions not hypothetical ones. For the full-blooded intentionalist, therefore, "the best historical explanation" is not constitutive of further intentions, but merely provides (defeasible) evidence of what those further intentions might have been.

How is this ambiguity to be resolved? Is Dyzenhaus's plain fact judge a full-blooded intentionalist who supplements law-making intentions with actual further intentions? Or is he rather an anaemic intentionalist with a Dworkinian twist, supplementing law-making intentions with hypothetical ones? There are good grounds, I think, for rejecting the latter characterisation of the plain fact judge. For there is no reason, as far as I can see, why a positivist who endorses the sources thesis because of the argument from moral intelligibility should accord hypothetical intentions (given content by a Dworkinian test of fit) any role in the determination of a community's law. It follows that, if Dyzenhaus's plain fact judge really did care for hypothetical rather than actual further intentions, if he really were an anaemic intentionalist with a twist of Dworkin rather than a full-blooded one, there would be no plausible way of connecting the plain fact approach to the sources thesis.

If, instead, Dyzenhaus's plain fact judge is taken to be a full-blooded intentionalist who concerns himself with further intentions actually possessed by the law-makers, the attempt to connect the plain fact approach to the sources thesis acquires some plausibility. A key notion in the argument from moral intelligibility is that the identifying criteria for a community's law must render intelligible participants' acceptance of law as authoritative. As we saw in our earlier discussion of the argument from moral intelligibility, this means that a community's law-makers must be able to make law in accordance with their moral judgement. Now, it may be felt that, in my presentation of the argument from moral intelligibility, I shied away from the full implications of the fact that law-makers must be able to make law in accordance with their moral judgement. From this fact, I drew the limited inference that the law of a community is to be identified by reference to social facts, as the sources thesis maintains.

[56] Dyzenhaus, *Hard Cases in Wicked Legal Systems* at pp. 56–7.

But, it may be said, it surely also follows that the law of a community is to be identified so as to reflect the intentions of the law-makers. For, it may be asked, if the law is not so identified, how will the law-makers be able to make law in accordance with their moral judgement? Once we have come this far, and accepted that the law must be identified so as to reflect the law-makers' intentions, it may seem that full-blooded intentionalism is irresistible. For, having once accepted the significance of the law-makers' intentions, what possible reason could there be to stop short of their further intentions?

The foregoing attempt to link acceptance of the sources thesis to acceptance also of full-blooded intentionalism is vulnerable to two objections. The first concerns its claim that, if a community's law-makers are to be able to make law in accordance with their moral judgement, the law they make must be identified so as to reflect their intentions. This claim seems to have the endorsement of Joseph Raz.[57] But I do not believe it to be correct. In order for the law-makers of a community to be able to express their judgement in the law, it is surely enough that the law be identified by the application to the law-makers' law-making acts of (linguistic) conventions that the law-makers knew, or could have known, were going to be applied to their law-making acts. If this is right, a commitment to the sources thesis, based on the argument from moral intelligibility, does not compel acceptance of intentionalism in any form.

I do not intend to pursue the first objection any further here. For, even if it were so that law-makers could make law in accordance with their judgement only if the law were identified so as to reflect their intentions, it would not follow that the intentions in question should include further intentions. In fact, and this is the second objection, quite the opposite is the case. If one accepts the sources thesis, and does so on the basis of the argument from moral intelligibility, one is specifically precluded from giving the law-makers' further intentions a role in one's identification of a community's law. For, paradoxically, if law were identified so as to reflect law-makers' further intentions, the ability of law-makers to make law in accordance with their moral judgement would be diminished rather than enhanced. Why so? Well, because the law-makers may well judge that their further intentions should not be reflected in the law. If, however, the law necessarily reflects the law-makers' further intentions, a judgement to this effect could not possibly be expressed in the law.

The idea that law-makers might judge it best to exclude their further intentions from any impact on the law they make is not as strange as it might at first appear. The following everyday example bears this out. Shortly before Christmas, I hand my daughter a sum of money, whilst uttering the words "Spend this well". My most immediate intention is a linguistic one. I intend to say just what these words ordinarily mean, given the conventions of the English language. But I have a further intention, namely that she spend the money buying gifts for others, rather than on herself. At the same time, I do not wish her

[57] Raz, "Intention in Interpretation", n. 36 above, at 256–68.

to buy gifts for others rather than treats for herself because of my further intention. For I wish her to act out of a generosity of spirit rather than out of compliance with my desires. In other words, I wish my further intention to be excluded from the reason(s) provided by my utterance.

Now the response may be made: "It may be so that you wish your further intention to be excluded. But behind your further intention (that your daughter buy gifts for others) is a further further intention, namely that she ignore your further intention. And you do not wish that to be excluded". But perhaps I do. Perhaps I wish my daughter to ignore my further intention, not because of my further further intention that she do so, but for reasons of her own. I wish her not only to be generous, but also to be a free spirit, an independent thinker. Of course, this in turn could be recast as a further further further intention that my daughter ignore my further further intention. But perhaps I wish this further further further intention to be excluded too. And perhaps I wish to exclude also any intention which might be lurking behind this further further further one.

The foregoing demonstrates a simple point. The law of a community does not only serve to provide the community's officials and members with reasons for action and decision. It also determines the extent to which the community's officials and members can act for reasons of their own. The law, in other words, has more than just a guiding function. It also has a jurisdictional one. It allocates original decision-making, decision-making on the basis of one's own reasons rather than reasons provided by others, between different public institutions and between public institutions and private individuals.

From this point follows another. If law has both a guiding and a jurisdictional function, then law can express the judgement of its makers in two respects: both in the reasons it provides and in the discretion it confers. A judgement in the latter regard will take account of a number of factors. One is the relative expertise of the law-makers and their subjects. Another is the limited ability of the law-makers to foresee changing social circumstances. But we need not concern ourselves with these factors here. My aim is to show that the argument from moral intelligibility rules out, rather than leads to, full-blooded intentionalism. For that, it is enough that we recognise the fact that law-makers have a judgement to exercise also as regards the jurisdictional aspect of the law they make.

Why so? Well, the argument from moral intelligibility requires that law-makers have the ability to express their judgement in the law they make. A community's law-makers will have the ability to express their judgement as regards law's jurisdictional aspect, only if they are able to place limits on the reasons they provide by their acts of law-making. A community's law-makers will lack this ability if, whenever they provide one reason by their law-making, one or more other reasons inevitably follow in train. Precisely this will be the case if, whenever law-makers perform a law-making act, some further intention (be it a proximate or a remote one) necessarily figures in the determination of the law they have made. Hence: full-blooded intentionalism is incompatible with the argument from moral intelligibility.

IV. LEGAL POSITIVISM'S UNDECLARED COMMITMENTS

According to David Dyzenhaus, a judge who endorses the positivists' sources thesis must also commit himself to the plain fact approach. Dyzenhaus also believes that adoption of the plain fact approach has morally undesirable consequences. Hence, concludes Dyzenhaus, legal positivism is invalid. I have argued, contra Dyzenhaus, that judicial endorsement of the sources thesis, founded on the argument from moral intelligibility, does not entail any commitment to the plain fact approach. For, as I have shown, nothing about the argument from moral intelligibility compels a judge who accepts it also to accept intentionalism or to regard law as authoritative (or, therefore, to deny that judges may exercise a discretion). If I am right in this, then Dyzenhaus's critique of legal positivism is simply a non-starter. Without a plausible connection between the plain fact approach and legal positivism, the iniquities of the former cannot in any way count against the merits of the latter.

Yet, although Dyzenhaus's critique of legal positivism is unsuccessful, it contains an important germ of truth. Dyzenhaus claims that acceptance of the sources thesis brings with it three further commitments, namely to intentionalism, to the authority of law and to an absence of discretion. I have shown this claim to be false. The claim is not false, however, because acceptance of the sources thesis brings with it no further commitments. It does, as I will explain below. The claim is false merely because the further commitments entailed by acceptance of the sources thesis do not include the three commitments identified by Dyzenhaus.

In this section I discuss two commitments that must, in my view, follow from acceptance of the sources thesis, at least when such acceptance is grounded on the argument from moral intelligibility. The first is that a community's law, to be law, must possess certain formal characteristics. The second is that some of a community's law, to be law, must emanate from institutions which possess certain formal characteristics. My claim that a positivist acceptance of the sources thesis entails these two further commitments clearly is revisionary of legal positivism. As we shall see, moreover, it is revisionary in a way which has an important implication for the way that legal positivism deals with a wicked legal system like the South African one under apartheid.

Let me start with the contention that acceptance of the sources thesis must lead to acceptance also of the view that law, to be law, must possess certain formal characteristics. Precisely what these formal characteristics are is not an issue that will be addressed in this essay. Suffice it to say that I have in mind characteristics roughly along the lines of Lon Fuller's "eight principles of legality".[58] These principles are sufficiently well-known not to require reiteration here.

[58] Fuller, *The Morality of Law*, n. 1 above, at pp. 33–94.

I should make it quite clear that I am not suggesting that acceptance of the sources thesis must lead to acceptance of all Fuller's claims regarding his eight principles of legality. Fuller makes three claims. One: conformity with the eight principles of legality is a necessary condition for law, i.e. if law deviates from these principles to a substantial degree, it ceases to be law.[59] Two: conformity with the eight principles of legality secures for the law a certain moral value. Three (this simply follows from one and two): law necessarily has moral value. The second of these claims has attracted a great deal of criticism. Hart, for example, has argued that respect for the principles of legality is "compatible with very great iniquity".[60] Conformity with these principles serves mainly to make the law more effective. But this is an efficiency which can be directed at evil as easily as at good. Although I believe Hart's criticism of Fuller's second claim to be well-founded, I will not be arguing the point here. For it is only Fuller's first claim, the claim that conformity with the principles of legality is a necessary condition for law, that is presently of interest. It is this claim that, in my view, is entailed by acceptance of the sources thesis.

Why does acceptance of the sources thesis, based on the argument from moral intelligibility, commit one also to accept formal requirements for law? Well, according to the argument from moral intelligibility, a proper understanding of law must render intelligible the beliefs of law's participants regarding it. This means that a proper understanding of law must render intelligible the belief that law is morally binding. The latter belief is intelligible only if it is at the very least possible that people do better by following the law than they do by following their own judgement. As I will now show, this possibility will not obtain unless the law has certain formal characteristics. Hence, law must be understood as possessing such formal characteristics.

Imagine a statutory provision that violates Fuller's principles of legality in any one of the following ways. It requires conflicting actions, it is ambiguous as to the action it requires, or it requires an action which it is impossible to perform. A statutory provision of this kind clearly is incapable of providing any guidance at all. Inevitably then, it cannot possibly provide a better guide to right action than do people's own judgements. As we have seen, the argument from moral intelligibility requires that law be understood in such a way that it *is* capable of improving on people's moral judgement. Thus, the argument from moral intelligibility requires that law be understood in a way which excludes from its ambit a statutory provision which violates Fuller's principles of legality in any one of the three ways described. Or, to put it another way, the argument from moral intelligibility requires that a directive be regarded as law only if it possesses the formal characteristics of consistency, clarity and possibility of performance.

A similar argument supports the contention that acceptance of the sources thesis must lead to acceptance also of the view that some of a community's law,

[59] Fuller, *The Morality of Law*, n. 1 above, at pp. 38–41, 197–200.

[60] Hart, *The Concept of Law*, n. 1 above, at p. 207; *Essays in Jurisprudence and Philosophy* (Oxford: Clarendon Press, 1983) at pp. 349–53.

to be law, must emanate from institutions which possess certain formal characteristics. Imagine a society in which everyone believes himself to be morally bound to follow the directives of a hermit who has had no contact with that society for several decades. Could a positivist who adheres to the sources thesis and does so because of the argument from moral intelligibility acknowledge the hermit's directives as law? The answer, surely, is that a positivist may be able to acknowledge some of the hermit's directives as law, but could not so regard all of them. The argument from moral intelligibility admits as law only that which can sensibly be believed to be morally binding. Thus, the argument from moral intelligibility admits as law only that which is capable of providing a better guide to right action than individual judgement. Though it may be possible that some of the hermit's directives improve on the moral judgements of those who live in the real world, it is not possible, in any realistic sense, that all do so.

It is instructive to consider exactly why not all of the hermit's directives could possibly improve on the judgement of people in the real world. For consideration of this question provides some insight into the formal features an institution must possess before a positivist adherent of the sources thesis can acknowledge all of its directives as law. The moral soundness of many directives by which a complex modern society is to be governed will not depend on moral principles alone. The moral soundness of some directives will depend on the preferences of society's members. As Ronald Dworkin and Joseph Raz have indicated, a choice will sometimes have to be made between directives, none of which has any intrinsic moral merit. In this event, the morally right directive is surely the one which best reflects popular preference.[61] It may also happen that a choice has to be made between directives of roughly equal intrinsic merit. Again, I would say, the morally right directive would have to be determined by popular preference. The moral soundness of many other directives will depend on the interests of society's members. Much can be said about interests and their moral relevance. But I will make only two points here. First: though people's preferences and interests will to a large degree coincide, they are not identical. The former at best provide defeasible evidence as to the latter. Secondly: while we possess some interests necessarily, simply by virtue of our being human, we possess many other interests contingently, because of choices we have made or circumstances we find ourselves in.

It should be clear, now, why not all of the directives issued by a hermit who has no contact with society can possibly improve on the judgements of society's members. The moral soundness of many actions will depend, wholly or in part, on preferences and contingent interests. Since the hermit has no way of identifying preferences and contingent interests, it is impossible that the hermit provide sound, let alone superior, moral guidance regarding these actions. It follows that some of the hermit's directives, namely those which depend for

[61] See Dworkin, "What is Equality? Part 4: Political Equality" (1987) 22 *Univ of San Francisco LR* 1 at 24; "Equality, Democracy, and Constitution: We the People in Court" (1990) 28 *Alberta LR* 324 at 331; Raz, "Liberalism, Skepticism, and Democracy" (1989) 74 *Iowa LR* 761 at 778–9.

their moral soundness on preferences and contingent interests, cannot be regarded as law by a positivist adherent of the sources thesis.

The example of the hermit suggests two formal characteristics that an institution must possess before all of its directives can be regarded as law by a positivist who adheres to the sources thesis on the basis of the argument from moral intelligibility. The first is that the institution must possess a mechanism for the identification of people's preferences and contingent interests. The second is that the institution must be such that, once people's preferences and contingent interests have been identified, they are accorded their proper weight. An institution which lacks these features might be able to issue some, perhaps even many, directives that a positivist will recognise as law. For it may be that some or many of the directives issued do not depend for their moral soundness on preferences and contingent interests. However, it is unavoidable that directives will be issued which do depend for their moral soundness on preferences and contingent interests. Directives of this kind, if issued by an institution lacking the described two features, will not be capable, in any realistic sense, of improving on the moral judgement of society's members. Since the argument from moral intelligibility supposes that law *is* capable of improving on individual members' moral judgements, anyone relying on the argument from moral intelligibility must deny that these directives are law.

If the two revisions of legal positivism that I have suggested are sound, then we will have to reconsider the relationship of legal positivism to a wicked legal system like that of apartheid South Africa. The usual understanding of this relationship, held both by positivism's champions and by its opponents, is as follows. One: legal positivism imposes no moral limits on law. Two: therefore legal positivism cannot deny that apartheid statutes were law. The revised positivism that I have set out does not in any way threaten the first of these propositions. But it seems to threaten the second. If the revised positivism that I have argued for is sound, then, though legal positivism might not set any moral requirements for law, it does set formal ones. And it appears that some of these formal requirements were not satisfied by the South African legal system during the apartheid years.

The following are uncontested facts about apartheid South Africa. Black South Africans could not be members of the South African Parliament. Black South Africans did not have freedom of political association. Black South Africans did not have freedom of political expression. Black South Africans did not have the vote. The cumulative effect of these and other facts is, I would suggest, that the South African Parliament had neither the means to identify nor the means properly to weigh the preferences and contingent interests of all South Africans. At the same time, it is undoubtedly so that the South African Parliament enacted many statutes the moral soundness of which depended on preferences and contingent interests. It was impossible, in any realistic sense, that these statutes would provide sound, let alone superior, moral guidance. This being so, a legal positivist who endorses the sources thesis, and does so

because of the argument from moral intelligibility, has good reason to deny that these statutes were law.

A positivist does not, of course, weaken his cause by making this denial. Does he strengthen it? Yes and no. Yes because, by making this denial, the positivist wrong-foots all those critics who decry positivism on the grounds that it recognises apartheid statutes as law. No because, as I explain in the conclusion to this essay, the validity of legal positivism is in no way affected by whether it does or does not recognise apartheid statutes as law.

V. LEGAL POSITIVISM AND THE RULE OF LAW

One of the difficulties in writing about the rule of law is that it is understood in a number of ways. Taken literally, the rule of law requires no more than fidelity to the law of one's community. A legislator, judge, policeman or citizen does all the rule of law requires of him, so long as he obeys his community's law. This is not, however, how most people understand the rule of law. Most people are of the view that, in order to observe the rule of law, a legislator, judge and so on has to conform also with a number of independent standards, standards that may be set down in a community's law but need not be.

Opinion is divided both over the nature of these independent rule of law standards and over the consequences of their infringement. Some legal philosophers, such as Lon Fuller, Joseph Raz and John Finnis, insist that these standards are entirely formal in nature.[62] The International Commission of Jurists, by contrast, has endorsed an understanding of the rule of law that is rather more substantive in content.[63] As far as the consequences of non-observance are concerned, the main issue seems to be whether non-compliance with the rule of law constitutes only a moral or also a legal failure. Those who regard non-compliance with the rule of law as a legal failure are not always clear about the meaning of this. One possibility is that they have in mind legal invalidity. That is, where a law-making institution acts in conflict with the requirements of the rule of law, it fails to make law. Or, where a law-applying institution fails to comply with the rule of law, it for that reason acts beyond its legal powers. It seems, however, that those who view non-compliance with the rule of law as constituting a legal failure often have something else in mind. Namely, regardless of whether an act which conflicts with the rule of law is for that reason legally invalid, it is in some other important sense legally defective.

Some critics of the South African judiciary's record during the apartheid era have sought to combine the view that the requirements of the rule of law are substantive in nature with the view that non-compliance with the requirements

[62] See Fuller, *The Morality of Law*, n. 1 above, at pp. at 96–7; Raz, *The Authority of Law*, n. 17 above, at pp. 210–19; Finnis, *Natural Law and Natural Rights*, n. 38 above, at pp. 270–6.

[63] See the discussion of this by Anthony Mathews, *Freedom, State Security and the Rule of Law* (Cape Town: Juta, 1986) at pp. 11–14.

of the rule of law results in a legal failure.[64] The attraction of this combination of views, I suspect, is as follows. Certainly, if the rule of law is given a substantive content, then the South African judiciary on many occasions failed to uphold the rule of law. If, moreover, non-compliance with the rule of law constitutes a legal failure, the South African judiciary can be twice condemned. They can be found guilty not only of moral but also of legal wrongdoing. Though this is not clear, it may be that those who advance this combination of views do so also because they assume that a legal failure on the part of the South African judiciary provides a more solid foundation for criticism than does a moral one.

Those critics of the South African judiciary who rely on the above understanding of the rule of law seem to think that this understanding of the rule of law is incompatible with legal positivism.[65] Indeed this is so. But I am not sure that the reasons for the incompatibility have been properly grasped. Nothing about legal positivism precludes its adherents from giving the rule of law a substantive content. Nor does anything about legal positivism stop an adherent from agreeing that non-compliance with the rule of law constitutes a legal, as opposed to a moral, failure. What a legal positivist cannot do, however, is to hold both of these views at once.

To see why this is so, we need to retrace our steps a little. In the previous section I argued that a positivist who endorses the sources thesis, and does so because of the argument from moral intelligibility, must also accept that law is formally constrained. Directives issued by a community's political institutions could fail to qualify as law either because the directives themselves lack certain formal features or because the institutions do. This is over-simplified. In the case of most formal characteristics, non-compliance may be of varying degrees.[66] A statute may be only slightly unclear or it may be greatly so. An institution may be slightly insensitive to preferences and interests, or it may take no account of them at all. This means that, if non-compliance with certain formal characteristics (such as clarity or sensitivity to preferences and interests) is to have the result that a directive fails to qualify as law, it cannot be non-compliance *per se*, but must be non-compliance of a particular degree. For no one could seriously suggest that any degree of non-compliance, even the most minor, must always result in a directive not being law.

If, in the case of some formal characteristic, it is only non-compliance of a particular degree that results in a directive not being law, there will be many directives which fail to comply with that formal characteristic, yet remain law. For there will be many directives which fail to comply with that formal charac-

[64] See, for example, Mathews, *Freedom, State Security and the Rule of Law*, n. 63 above, at pp. xxix, 15–22; Dyzenhaus, *Hard Cases in Wicked Legal Systems*, n. 4 above; "Law's Potential" (1992) 7 *Canadian Journal of Legal Studies* 237.

[65] See Mathews, *Freedom, State Security and the Rule of Law*, n. 63 above, at pp. 297–302; Dyzenhaus, *Hard Cases in Wicked Legal Systems*, n. 4 above; "Law's Potential", n. 64 above.

[66] See Raz, *The Authority of Law*, n. 17 above, at p. 215.

teristic, but not to the required degree. Thus a statute may be unclear, but not so unclear that it fails to qualify as law. Or a statute may be issued by a political institution which is insensitive to preferences and interests, but not so insensitive that it cannot be regarded as law-making. How are directives of this kind to be regarded? More specifically, how are they to be regarded by a positivist who endorses the sources thesis, and does so because of the argument from moral intelligibility? *Ex hypothesi*, these directives cannot be regarded as non-law. But should they perhaps be regarded as legally defective in some other important sense?

In my view, a positivist who endorses the sources thesis because of the argument from moral intelligibility has good reason to answer the latter question in the affirmative. The argument from moral intelligibility involves a functional understanding of law, namely that the primary aim of law is to provide its subjects with a superior guide to the moral reasons that apply to them. Now, where a kind is functionally defined, the following distinction can typically be drawn. Some objects will simply be incapable of performing the relevant function. These objects are thus not of that kind. Other objects are capable of performing the relevant function, but less so than is possible. These objects certainly are of that kind, but are poor or defective instances thereof. Joseph Raz provides an illustration of this point.[67] A knife's function is to cut. Thus an object is not a knife if it cannot cut at all. Thus also an object is a deficient knife if it can cut, but cuts badly because it is blunt.

We have just seen that, in the case of some formal characteristics, it is not non-compliance *per se*, but only non-compliance to a certain degree, which results in a directive not being law. The reason for this, we now know, is because law's function is to improve on its subjects' moral judgements and because non-compliance to that degree altogether deprives a directive of its ability to provide such improvement. A directive which fails to comply, but to a lesser degree, has the ability to improve on subjects' judgements. It is thus able to discharge law's function and is therefore law. However, to the extent that such a directive fails to comply, it is less able to improve on subjects' judgements than it would have been, had it complied. The directive thus suffers from an impaired ability to perform law's function. This means that, though the directive is law, it is less good as law than it could have been. Or to put it another way, though the directive is legally valid, it is legally defective.

The foregoing discussion establishes two things. Most obviously, it shows that the understanding of law which underpins the positivists' sources thesis generates not merely criteria of legal validity and invalidity, but also criteria of legal merit and deficiency. Less obviously, it suggests a limit upon the criteria which the positivist understanding of law can generate. For the line of reasoning which motivates these criteria establishes formal criteria only and not substantive ones. Why so? Because, as we have seen, the argument from moral

[67] Raz, *The Authority of Law*, n. 17 above, at pp. 225–6.

intelligibility involves a functional view of law. Law is meant to improve on the moral judgements of its subjects. As we have also seen, this functional understanding of law generates certain formal criteria of legal validity and merit. There would appear to be no way, however, that such a functional understanding of law can establish substantive criteria of legal validity and merit.

This has an important implication for the relationship of legal positivism to the rule of law. In particular, it explains why a legal positivist who accepts the sources thesis, and does so because of the argument from moral intelligibility, cannot both regard the rule of law as a legal (rather than a moral) doctrine and give it a substantive (rather than a formal) content. Ask yourself the question: what could make it the case that the rule of law is a legal rather than a moral doctrine? Only one answer is plausible, namely: something about the nature of law. I have just argued that the understanding of law's nature which underlies the positivists' sources thesis generates only formal criteria for legal validity and merit, not substantive ones. If this is correct, then, to the extent that the rule of law is a legal doctrine, it cannot be substantive in content. At least, it cannot be substantive in content if you are a legal positivist who endorses the sources thesis because of the argument from moral intelligibility.

I should emphasise that I have not shown that a positivist who endorses the sources thesis, and does so because of the argument from moral intelligibility, cannot give the rule of law a substantive content. I have only shown that a positivist cannot do this and also maintain that the rule of law is a legal (as opposed to a moral) doctrine, in the sense that non-compliance with the rule of law constitutes a legal (rather than a moral) failure. But, if it is so that a positivist cannot do both these things at once, then there is good reason for a positivist to resist the temptation of the former. That is, the positivist should refrain from giving the rule of law a substantive content. For the power of the rule of law doctrine, and of criticism directed at those who disregard its precepts, is to a large degree dependent upon its being a legal (rather than moral) doctrine. It follows that, to the extent that a legal positivist were to give the rule of law a substantive content, he would be depriving the rule of law of its significance and usefulness. The more substantive the rule of law becomes, the less reason a positivist has to pay it special attention.

Now, as was mentioned earlier, some critics of the South African judiciary's record under apartheid assume both that the rule of law has a substantive content and that it is a legal doctrine, in the sense that its violation invites legal reproach. If legal positivism (in the guise of the sources thesis as justified by the argument from moral intelligibility) is valid, then these critics have a difficulty. Either they have to relinquish the rule of law's substantive content: in this event the South African judiciary violated the rule of law a great deal less than these critics would have us believe. Or they have to forego the idea that the rule of law is a legal doctrine: in this event the South African judiciary cannot be twice condemned, for failing both morally and legally. Either way, the criticism of the South African judiciary's record loses some of its sting.

Given this implication of legal positivism for their critique of the South African judiciary, one would expect these critics to attempt a refutation of legal positivism. Indeed, one is not disappointed.[68] However, the refutation of positivism proceeds in a surprising fashion. One: as shown by the context of South Africa under apartheid, it is desirable that the rule of law have substantive content and that it be regarded as a legal doctrine. Two: the rule of law can combine these two features only if legal positivism is invalid. Three: therefore legal positivism is invalid. In other words, rather than argue that positivism is false and therefore does not threaten their conception of the rule of law, these critics argue that positivism is false because it threatens their conception of the rule of law.

Once again, we have hit jurisprudential bedrock. Undergirding these critics' way of dealing with the threat positivism poses to their understanding of the rule of law is an assumption that we have come across before, namely that the validity of a legal theory depends upon the desirability of its practical effects. As we have seen, this assumption is crucial to Dyzenhaus's attack on legal positivism. The assumption is also vital to the idea that the validity of legal positivism might be affected by whether it recognises apartheid statutes as law. Now I have clearly expressed my rejection of this assumption. But I have as yet provided no reasons for this rejection. The conclusion that follows goes some way towards remedying this.

CONCLUSION

In this conclusion I wish to provide some support for my rejection of the idea that the acceptability or validity of a legal theory is dependent upon the desirability of its practical effects. Let us call this the "desirable effects" condition. And let us start our argument against this condition by recalling the first three points in the argument from moral intelligibility. According to point one, a legal theory is acceptable only if it is sensitive to the beliefs of law's participants about it. Point two identified, as one of these beliefs, the belief that law is morally binding, that it has legitimate authority. And point three claimed that a theory of law is sensitive to this belief as long as it renders the belief morally intelligible. It need not, however, render the belief true.

Now it is possible that those who endorse the desirable effects condition mean to deny (or qualify) point one in the argument from moral intelligibility. That is, they mean to say that sensitivity to the beliefs of law's participants is not a (or the) standard by which the acceptability of a legal theory is to be measured. If so, they are the minority defenders of a wholly implausible view. On the one hand, and as indicated in the earlier discussion of the argument from moral intelligibility, point one enjoys almost universal acceptance among contemporary legal

[68] See Mathews, *Freedom, State Security and the Rule of Law* at, n. 63 above, at pp. 297–302; Dyzenhaus, *Hard Cases in Wicked Legal Systems*, n. 4 above; "Law's Potential", n. 64 above.

philosophers. On the other hand, given that law is a social practice in large part constituted by the beliefs of its participants, a theory of law which fails to be sensitive to those beliefs has surely lost touch with the very phenomenon to which it supposedly relates.

There is, however, another possibility. This is that those who claim that the acceptability of a legal theory is conditional upon the moral desirability of its effects mean only to deny point three in the argument from moral intelligibility. That is, they agree that an acceptable theory of law must be sensitive to the beliefs of law's participants. But they reject the idea that, for a theory of law to be sensitive to such beliefs, it is enough that the theory render those beliefs intelligible. They believe, instead, that an acceptable theory of law must render those beliefs true. Of course, if this is correct, then it does follow that a theory of law is unacceptable if it has undesirable moral consequences. For, as point two in the argument from moral intelligibility makes plain, one of the beliefs that law's participants have is that law is morally binding, that law has legitimate authority. For this belief to be true, rather than merely intelligible, it presumably does have to be the case that the law is by and large good rather than wicked.

If this is the view of those who endorse the desirable effects condition, then they may be able to claim an ally in Ronald Dworkin. For Dworkin's assertion that a social scientist wishing to understand a practice like law should "*join* the practice he proposes to understand" and should "participate in the spirit of its ordinary participants" suggests that an acceptable theory of law does more than merely render the beliefs of law's participants intelligible.[69] Is there, then, any reason to prefer the view that an acceptable theory of law should merely render the beliefs of law's participants intelligible? Here is the beginning of an answer. Law is essentially institutional in nature. Any theory of law must take account of this fact. To be institutional is to be humanly conditioned. To be human is to be morally fallible. Hence, a necessary consequence of law's institutional nature is that it is morally fallible. But if law inevitably is morally fallible, then we can require no more of an acceptable theory of law than that it render the belief that law is morally binding intelligible. To require more, to require that an acceptable legal theory render this belief true, is to deny the moral fallibility of law.

[69] See Dworkin, *Law's Empire*, n. 8 above, at pp. 64, 422 n. 14.

6

Legal Positivism and American Slave Law: The Case of Chief Justice Shaw

ANTHONY J. SEBOK*

1. THE CONTESTED HISTORY OF LEMUEL SHAW

Perhaps no judge symbolises the crisis in American law during the era of slavery better than Lemuel Shaw. Shaw, who was Chief Justice of the Massachusetts Supreme Judicial Court from 1830 to 1860, was the most prominent jurist in the most anti-slavery state in the North. Shaw was a committed abolitionist and one of the nineteenth century's ablest judges, yet by 1851, after he declared that the federal government and not Massachusetts had exclusive power over the treatment of fugitive slaves, he was vilified as a traitor and slaver. Shaw symbolised a crisis in America's relationship with its new constitution. If Shaw was right, then the Constitution not only permitted the evil of slavery in the South, but extended the assumptions of slavery into the North as well.

The Constitution was amended to forbid slavery in 1865, but Shaw's fugitive slave decisions still provoke debate. This is as it should be. The question of whether Shaw's conclusions were legally compelled, or even legally sound, matters to anyone who cares about the relationship between adjudication and evil law. In recent years, the debate concerning Shaw's decisions has taken a curious turn. Rather than debate whether Shaw's conclusions were right, contemporary scholars assume that Shaw was wrong, and then move directly to the question of explaining *why* a great judge went so wrong. On this question there are two leading schools of thought. The best known argument is by Robert Cover. He argued that the fugitive slave cases posed a stark dilemma for Shaw between following the letter of the law and doing the right thing, and that Shaw resolved that dilemma by deciding those cases "formalistically": that is, he tried to minimise the role his own judgement and values played in the interpretation of the law by treating the law as fixed and autonomous, and his decision as inexorable or preordained.[1] William Nelson, however, has argued just the opposite claim.

* Thanks to John Goldberg, Susan Herman, G. Edward White, Steve Winter, and Benjamin Zipursky for comments and suggestions. Margaret Foley, BLS Class of 1999, provided invaluable research assistance.

[1] See Robert M. Cover, *Justice Accused* (New Haven: Yale University Press, 1975), pp. 4, 250–2.

According to Nelson, Shaw decided the fugitive cases "instrumentally": that is, he ignored the language of the law as well as the principles which lay behind the law in order to reach decisions that he thought would further the goal of preventing a war between the North and South.[2]

The disagreement between Cover and Nelson presents an interesting problem. Obviously, both scholars read the same facts differently, but, more importantly, they offer very different diagnoses for how the crisis created by Shaw's decisions could have been avoided. Implicit in both critiques is the view that Shaw was prevented from more just outcomes by paying too much attention to the wrong kind of legal theory: formalist positivism in Cover's view and instrumentalism in Nelson's view. As I will argue, both Cover and Nelson misinterpret Shaw's jurisprudence and its consequences. I will show that Shaw did not deploy formalism or instrumentalism in deciding the fugitive slave cases, but instead adopted a subtle form of legal positivism. The key to a better understanding of Shaw resides in taking cognisance of Shaw's *reasoning* in earlier slave law cases; reasoning that Cover and Nelson ignore.

II. SLAVES IN TRANSIT AND THE SOMERSET CASE

Northern judges confronted slavery in two types of cases. The first involved slaves who were brought voluntarily by their masters to a free state, and the second involved slaves who escaped from a slave to a free state. The first category can be referred to as "transit" cases while the second category can be referred to as "fugitive" cases.[3] The transit cases in turn fell into two periods. In the first period, which took place between 1787 and the mid-1830s, Northern judges extended "comity" to Southern slaveholders who entered Northern states with their "property". Under the comity approach, a free state recognised (or tolerated) a Southerner's claim to property in an enslaved person brought by the Southerner into the free state. In 1787 the only free state that explicitly modified the comity extended to visiting slaveholders was Pennsylvania, which, by statute, restricted the slaveholder's enjoyment of a right to property in persons for no longer than six months.[4] In 1801 New York limited the period of comity for slaveholders who sojourned in New York to nine months.[5] New Jersey, Rhode Island, Connecticut, Massachusetts, Ohio, Indiana, and Illinois

[2] See William E. Nelson, "The Impact of the Antislavery Movement Upon Styles of Judicial Reasoning in Nineteenth Century America", (1974) 87 *Harvard Law Review* 513 at 540 (hereinafter "Impact").

[3] For a good survey of the transit cases, see Paul Finkelman, An Imperfect Union: Slavery, Federalism, and Comity (Chapel Hill: University of North Carolina Press, 1981) and Aviam Soifer, "Compromise at the Boundaries of Bondage", (1982) 10 *Review of American History* 185; for a good survey of the fugitive cases, see Cover, *Justice Accused*, n. 1 above; and Donald Fehrenbacher, *Slavery, Law and Politics* (New York: Oxford University Press, 1981).

[4] See Finkelman, *An Imperfect Union*, n. 3 above, at pp. 49–55.

[5] See ibid., p. 72.

extended full comity until the 1830s to 1840s.[6] It is important to note that Southern states extended full comity to blacks freed by the operation of Northern law as well.[7]

Given conventional choice of law rules in the late eighteenth and early nineteenth century, the extension of comity by the Northern states to Southern slaveholders would not have been remarkable were it not for the fact that it was completely and utterly in contradiction with settled common law principles. As Story noted in his treatise on conflicts of laws, although states could expect foreign states to extend comity in cases involving real property or contract, "no nation [was] under any obligation to give effect to the laws of another nation, which [were] prejudicial to itself or its citizens", including anything "which [was] injurious to their public rights, or offend[ed] their morals, or contravene[d] their policy, or violate[d] a public law".[8] Moreover, the single most famous application of the public policy exception to comity occurred in a 1772 slavery case, *Somerset v Stewart*.[9] Somerset, a slave in Virginia, was brought by his master to England. When Somerset attempted to flee, his master imprisoned him in a ship in the Thames River. Somerset sued, and Lord Mansfield ordered Somerset free. Of course the slaveholder argued that, since slavery was legal in Virginia (a colony with its own domestic laws), English courts should have extended comity towards a Virginian sojourning in England. Mansfield flatly rejected this argument. Americans who read of the case found this reply: "the state of slavery is of such a nature, that it is incapable of being introduced on any reasons . . . but only by positive law . . . It is so odious, that nothing can be suffered to support it but positive law".[10] Although comity may be the norm when dealing with topics in which variation in municipal law might be expected (and even encouraged), such as contract law or property law, slavery stood on another footing altogether. It was so clearly in conflict with natural law that it could only come about by an explicit and deliberate command of the sovereign. It is worth noting that built into Mansfield's argument (which followed Blackstone on this point) was the principle that the *common law* could not support slavery, only legislation could.[11]

[6] See ibid., pp. 76–98.

[7] See Jacob D. Wheeler (ed.), *A Practical Treatise on the Law of Slavery* ("Of the Emancipation of Slaves By the Effect of Foreign Laws") (1837, reprinted New York: Negro Universities Press, 1968, pp. 334–56); Finkelman, *An Imperfect Union*, n. 3 above, at pp. 234–5; Louise Weinberg, "Methodological Interventions and the Slavery Cases; or, Night-Thoughts of a Legal Realist" (1997), 56 *Maryland Law Review* 1316 at 1332–5 and see infra text accompanying n. 101.

[8] Joseph Story, *Commentaries on the Conflict of Laws*, 1st edn. (Boston: C.C. Little & J. Brown, 1834), p. 30.

[9] *Somerset v Stewart*, (1772) 98 Eng. Rep. 499 (KB) and see Paul Finkelman, " 'Let Justice Be Done, Though the Heavens May Fall': The Law of Freedom", (1994) 70 *Chicago—Kent Law Review* 325 at 325–7 (on *Somerset*).

[10] As reported at 1 Lofft's Rep. 1 and 20 Howell's State Trials 1. As Cover noted, recent scholarship suggests that the version of *Somerset* which reached America through Lofft's Reports may not have been entirely accurate. Cover, *Justice Accused*, n. 1 above, at p. 17 n. 29.

[11] As Cover noted, Mansfield decided *Somerset* by using natural law as a state of affairs against which legislation was measured. Where there was no positive law on a matter controlled by natural

Somerset was well known to American lawyers and there is some evidence to suggest that it was in the minds of at least some of the delegates to the Constitutional Convention.[12] Nonetheless, Northern states still ignored it by extending comity to slaveholders. As Finkelman has demonstrated, despite half-hearted attempts to raise the question of slaveholder's rights in transit, the delegates to the Constitutional Convention focused their attention exclusively on the right of slaveholders to recapture fugitive slaves. This narrow focus on fugitives is both odd and significant given that the rule in *Somerset* did not distinguish between the means by which a slave reached English shores: Mansfield's holding did not limit itself to *fugitive* slaves, although Somerset indeed had been a fugitive. It is possible but unlikely that Southern delegates thought that the Full Faith and Credit and the Privilege and Immunities Clauses would protect their rights while visiting Northern states. Debate over the Full Faith and Credit Clause was brief and limited to the questions of bankruptcies and foreign bills of exchange.[13] Debate over the Privileges and Immunities Clause did elicit an objection from Charles Pinckney of South Carolina, who said that "some provision should be included in favor of property in slaves".[14] Yet as Finkelman pointed out, Pinckney's concerns were ignored and did not create a controversy in a Convention in which slavery was jealously protected by the South.[15] The most likely explanation for the absence of controversy over the rights of slaveholders in transit is that the South did not feel a need to protect those rights: the Northern states, *Somerset* notwithstanding, had given every indication that they would extend comity to slaveholders in transit.[16] Whatever the reason, between 1787 and 1836 *Somerset* was mostly ignored by American courts.[17]

In contrast to the first, the second period of "transit" cases saw the Northern courts do an about-face and embrace *Somerset*. In fact, by 1860 every Northern

law, the common law was presumed to have adopted natural law. But where there was positive law in conflict with natural law, the judge was obliged to follow positive law. See Cover, *Justice Accused*, n. 1 above, at p. 17.

[12] See David Brion Davis, *The Problem of Slavery in the Age of Revolution* 1770–1823 (Ithaca: Cornell University Press, 1975), 278; William M. Wiecek, *The Sources of Antislavery Constitutionalism in America*, 1760–1848 40–41 (Ithaca: Cornell University Press, 1977), pp. 40–1; William M. Wiecek, "Somerset: Lord Mansfield and the Legitimacy of Slavery in the Anglo-American World", (1874) 42 *University of Chicago Law Review* 87, 117–18; Jerome Nadelhaft, "The Somersett Case and Slavery: Myth, Reality, and Repercussions", (1966) 51 *Journal of Negro History* 193, 202.

[13] See Finkelman, *An Imperfect Union*, n. 3 above, at p. 33 (citing Max Farrand (ed.), *The Records of the Federal Convention of 1787* (1937), vol. 2, p. 447).

[14] Finkelman, *An Imperfect Union*, n. 3 above, at p. 35 (citing Max Farrand (ed.), *The Records of the Federal Convention of 1787* (1937), vol. 2, p. 443).

[15] Finkelman, *An Imperfect Union*, n. 3 above, at p. 35.

[16] Ibid., at p. 40.

[17] Only a handful of cases considered *Somerset*. See *Pirate, alias Belt v Dalby*, 1 Dall. 167 (Pa. 1786) (rejecting *Somerset*); *Mahoney v Ashton*, 4 H. & McH. 295 (Md. 1799) (rejecting *Somerset*); and *Denison, et. al. v Tucker*, 1 Blume Sup. Ct. Trans. 385 (Mich. 1807) reprinted in (1935) 1 Transactions of the Supreme Court of the Territory of Michigan (suggesting endorsement of *Somerset*).

state had adopted *Somerset*.[18] And the leader of this change in course was none other than Lemuel Shaw.

Shaw, who had become Chief Justice in 1830, first indicated his impatience with the extension of comity to slaveholders in 1832 in the case of *Commonwealth v Howard*.[19] There, the court was asked to free Francisco, a young boy of twelve or fourteen who had been brought to Massachusetts as a slave from Cuba by his master, Mrs Howard.[20] Mrs Howard responded that Francisco was not her slave and was accompanying her to Cuba voluntarily, as her servant. Shaw interviewed Francisco in private and concluded that the boy freely wished to accompany Mrs Howard. Shaw therefore dismissed the case. In his decision, which was not officially reported but was summarised in the journal the *Daily Atlas*, Shaw stated that "if Mrs. Howard, in her return to the writ, had claimed the boy as a slave, I should have ordered him to be discharged from her custody".[21]

In 1836 Shaw was confronted with a pure "transit comity" question. Med, a six-year-old black girl, was brought to Boston by Mrs Mary Slater from Louisiana. Med was a slave owned by Slater's husband. While in Boston, Slater and Med lived with Slater's father, Thomas Aves. The Boston Female Anti-Slavery Society, as well as other abolitionists, sued under a writ of habeas corpus to have Med freed. Benjamin Curtis, the lawyer for Aves, argued that *Somerset*, which had been raised by Med's lawyers, was inapplicable by reason of what Levy called the "peculiar and intimate relationship among sister American states".[22] Shaw rejected Curtis's argument and adopted *Somerset* in its entirety: "[common law] decides that slavery, being odious and against natural right, cannot exist, except by force of positive law. But it clearly admits that it may exist by force of positive law. And it may be remarked, that by positive law in this connection, may be as well understood customary law as the enactment of a statute".[23] Since Massachusetts abolished slavery in 1780, there were neither statutes nor customary law supporting slavery in Massachusetts in 1836. Thus, Shaw adopted Mansfield's argument that slave law was simply different from other municipal laws concerning contract or property; and Massachusetts could ignore the positive law of Louisiana if it wanted. But Shaw reached further and made a more subtle argument. Since nature abhors slavery, a person could not be *property*, even in Louisiana. Louisiana could "for its own convenience, declare that slaves be *deemed* property, and that the relations and laws

[18] Some states adopted *Somerset* by judicial holding, others by legislation. See Finkelman, *An Imperfect Union*, n. 3 above, at pp. 127–80.

[19] 18 Pick. 193 (Mass. 1836).

[20] This case was discussed by various journals. See, e.g., *The Liberator*, 8 December 1832, and (1832) 9 *The American Jurist and Law Magazine* 490.

[21] Leonard Levy, *The Law of the Commonwealth and Chief Justice Shaw* (Cambridge: Harvard University Press, 1957), p. 62 (hereinafter *Law of the Commonwealth*) (quoting original sources including the *Daily Atlas*).

[22] Levy, *Law of the Commonwealth*, ibid., at p. 64.

[23] *Commonwealth v Aves*, in Jacob D. Wheeler (ed.), *A Practical Treatise on the Law of Slavery* (1837), p. 364.

of personal chattels shall be *deemed* to apply to them . . . but it would be a perversion of terms to say, that such local laws do in fact make them personal property".[24]

Shaw's move here is very interesting: he argued that slavery was not an aspect of property law but of criminal and tort law. The law of slavery in Louisiana did not create a new form of property but created a new set of criminal immunities and tort privileges: the reason why Mr Slater could command Med was not that Med was property, but because Mr Slater could not be prosecuted for assaulting Med and Med could not sue Mr Slater in tort. Thus, Shaw concluded, "as a general rule, all persons coming within the limits of a state, become subject to all its municipal law, civil and criminal, and entitled to the privileges which those laws confer; that this rule applies to blacks as whites . . . that if such persons have been slaves, they become free, *not so much because any alteration has been made in their* status, *or condition, as because there is no law which will warrant . . . their forcible detention or forcible removal*".[25]

It has been noted that Shaw justified his adoption of *Somerset* on the grounds that the extension of comity to visiting slaveholders would "extend [slavery] to every place where slaves might be carried".[26] While this is true, it is important to recognise that it was neither the practical nor theoretical threat of Southern slaveholders invading the North that motivated Shaw's decision. Rather, Shaw viewed the recognition of even a single slaveholder's claim not as a demand that Massachusetts borrow from Louisiana's property law, but as a demand that Massachusetts distort its entire body of criminal and tort law. Just as slavery was a wholesale reconfiguration of the criminal and tort law of Louisiana, the extension of comity to Mr Slater would require the wholesale reconfiguration of Massachusetts' criminal and tort law. That is why Shaw rejected the extension of comity to a visiting slaveholder: unlike a visitor who wanted Massachusetts to recognise a species of real property unique to his home state, the slaveholder asked that Massachusetts do something which was "*wholly* repugnant to [Massachusetts's] law, *entirely* inconsistent with [Massachusetts's] policy".[27]

The distinction between comity for a discrete concept of property and comity for a complex network of criminal and tort immunities may explain why Shaw felt the need to devote almost a third of *Aves* to the meaning of the Fugitive Slave Clause of the Federal Constitution. It is not surprising that Curtis would have tried to argue that the clause, which prohibited any state from using its law to free a fugitive slave, also required a state to extend comity to visiting slaveholders. What is surprising is that, instead of disposing of the argument with a quick textual reference to the fact that the clause dealt with *fugitive* slaves, Shaw set out a long explanation of the history and purpose of the clause. The clause, he noted, was designed to solve a problem that inevitably followed from *Somerset*.

[24] *A Practical Treatise on the Law of Slavery* (1837), p. 368 (emphasis added).
[25] Ibid., at p. 369 (emphasis added).
[26] Levy, *Law of the Commonwealth*, n. 21 above, at pp. 65–6.
[27] *Aves*, n. 23 above, at 369 (emphasis added).

Before the Constitution (ostensibly during the years of the Articles of Confederation) the states related to each other in the same way that England related to other countries: as "sovereign and independent" nations.[28] While independent nations usually have no trouble (or no special trouble) extending comity to each other's citizens through the rules of private international law, slavery would have posed a special problem for the reasons explained above. Thus, if the states had continued as sovereign nations until 1836, Shaw argued, it would be likely that the states would have set up treaties between them to handle the inevitable controversy that would attend the escape of a slave from a Southern to a Northern state.[29] But, because of the profound effect of the practice of slavery on a state's law, these treaties would have been the subject of intense negotiations between the two parties. Shaw hypothesised that it would have been the "the intent and the object of one party to this compact [the Southern state] to enlarge, extend and secure, as far as possible, the rights and powers of the owners of slaves, within their own limits, as well as in other states, and of the other party [the Northern state] to limit and restrain them".[30]

From this thought experiment Shaw concluded that any "agreement" between the free and slave states would have had the following features. (1) The slave states would have insisted on retaining "plenary power to make all laws necessary for the regulation of slavery and the rights of slave owners, whilst the slave remain within their territorial limits".[31] (2) The free states would have insisted that only when slaves escaped into other states would the slave states be able to demand "the aid of other states to regain their dominion over the fugitives".[32] The Fugitive Slave Clause was the best bargain the South could get if negotiations began with *Somerset* as the baseline. The Clause's effect was thus to "limit and restrain the operations of" *Somerset* by suspending its scope in the case of "fugitive" slaves—slaves entering Northern territory *against* their master's wishes.[33] Shaw went into such detail about the Fugitive Slave Clause because, if it indeed was negotiated in the shadow of *Somerset*, then it was extremely good evidence of the fact that *Somerset* had been and still was good law in Massachusetts (and, by extension, all of the free states).

Shaw was praised throughout the North for his decision. The leading abolitionist William Lloyd Garrison called *Aves* the "rational, just and noble decision of [an] eminent judge".[34] The Boston *Columbian Centinel* called the decision "the MOST IMPORTANT" ever made in any of the free states.[35] The reaction in the South was "divided only in its degree of disapprobation" although one

[28] Ibid., at p. 370.

[29] Shaw reasoned that "such a stipulation would be highly important and necessary to secure peace and harmony between adjoining nations and to prevent perpetual collisions and border wars": ibid., at p. 371.

[30] Ibid..

[31] Ibid., at p. 372.

[32] Ibid.

[33] Ibid., at p. 370.

[34] *The Liberator*, 24 Sepember 1836.

[35] Quoted in Levy, *Law of the Commonwealth*, n. 21 above, at p. 67.

paper, the Louisville *Advertiser*, did not see the decision as a great threat to Southern interests.[36] Some Southerners seemed to think that by recognising *Somerset* and rejecting comity, Shaw had "annulled" the Fugitive Slave Clause.[37] Of course, nothing could have been further from the truth.

Shaw built on *Aves* in his decision in *Commonwealth v Porterfield*, in which he freed a slave who had been brought to Boston from New Orleans en route to Cuba (a slave jurisdiction) on a brig.[38] Although the slaveholder had not "wanted" to come to Boston (his goal was to go to Cuba), Shaw held that the slave was not a fugitive and no law could restrain him in Boston; he could leave the brig and remain in Massachusetts if he chose (which is what happened). In *Commonwealth v Fitzgerald*, Shaw freed a slave named Robert Lucas, who had been brought to Boston on a Navy frigate after a two year voyage that had originated in Virginia. The slaveholder argued that he had not brought Lucas "voluntarily" into Massachusetts, since the Navy had ordered the ship to Boston.[39] Shaw rejected this argument, too: he stated that since the slaveholder had "consented that the slave should be carried anywhere that the ship might be sent", Lucas was not a fugitive and could not be held against his will on the frigate.[40]

Aves soon became established law in the North.[41] Some major state courts soon adopted *Aves*: Connecticut in 1837; Ohio in 1841; arguably New York in 1846; and Pennsylvania in 1849.[42] Other states recognised the principle underlying *Aves* through legislation or state constitutional provisions. As Finkelman put it, "By 1860 transit with slave property had no protection in most of the North. Fifty years of comity disappeared in less than half that time".[43] In short, Shaw had enacted a legal revolution that was every bit as influential as his reorganisation of tort law around negligence in *Brown v Kendall*.[44]

[36] Quoted in Levy, *Law of the Commonwealth*, n. 21 above, at at p. 68.

[37] Baltimore *Chronicle*, quoted in Levy, *Law of the Commonwealth*, n. 21 above, at p. 68.

[38] As reported in 7 *Monthly Law Reporter* 256 (Mass. 1844).

[39] As reported in 7 *Monthly Law Reporter* 379 (Mass. 1844).

[40] Ibid., at p. 382.

[41] Story reproduced the decision in its entirety in an extended footnote covering 17 pages in the next edition of his treatise on the conflict of laws. See Joseph Story, *Commentaries on the Conflict of Laws*, 2nd edn. (Boston: C.C. Little & J. Brown, 1841).

[42] *Jackson v Bulloch*, 12 Conn. 38 (1837); *State v Farr* (Ohio 1841) (unreported case discussed in *Niles Weekly Register*, 29 May 1841); *In re George Kirk*, 1 Parker Cr. R. 67 (N.Y. 1846); and *In re Lewis Pierce*, 1 Western Legal Observer 14 (Pa. 1849).

[43] Finkelman, *An Imperfect Union*, n. 3 above, at p. 127. Soon after *Aves*, Southern states began to gradually refuse the effects of Northern laws on blacks travelling in the South, and by 1860 almost no Southern state extended comity to blacks freed by the operation of Northern law. See Finkelman, *An Imperfect Union*, n. 3 above, at p. 234 but see Weinberg, "Methodological Interventions and the Slavery Cases", (1997) n. 7 above, at pp. 1338–43 (although ultimately the South abandoned comity, it recognised Northern laws that freed blacks for longer than Finkelman claimed).

[44] 6 Cush. 292 (Mass. 1850) (adopting fault standard and rejecting strict liability in accident cases). See Charles Gregory, "Trespass to Negligence to Absolute Liability", 37 *Virginia Law Review* 361at 365–9 (Shaw "gets most of the credit for the establishment of a consistent theory of liability for unintentionally caused harm") and G. Edward White, *Tort Law in America* (New York: Oxford University Press, 1980), p. 15 ("*Brown v Kendall's* significance lay in Shaw's recognition of the capacity of 'fault' to serve as a comprehensive standard").

III. FUGITIVE SLAVES AND THE LIMITS OF *SOMERSET*

Shaw's decisions in "transit" cases like *Aves* were soon to be overshadowed by his decisions in the fugitive slave cases. Shaw's first encounter with the Fugitive Slave Clause suggested that he wished to avoid applying it if possible. In 1836 two passengers on the brig *Chickasaw*, Eliza Small and Polly Ann Bates, were seized in Boston harbour by the ship's captain on behalf of a slave-catcher who claimed that the women were fugitives from a slaveholder in Baltimore.[45] The Fugitive Slave Law of 1793 required that a fugitive slave could be held and returned to the state from which they had fled only after a finding by a federal judge or a state magistrate, a condition the captain had failed to meet. As reported in newspaper accounts, a writ of habeas corpus was obtained, and the captain was forced to appear before Shaw to explain why his detention of the women was lawful.[46] After the case had been argued, Shaw bluntly concluded that the ship's captain's detention of the women was illegal, since it was not based on any prior judicial determination.[47] Of course, it might have been possible for Shaw himself to make the necessary finding under the 1793 law, but for that to be done the women would have to be rearrested. The slave-catcher on whose behalf the captain had acted rose to state that he intended to make fresh arrest under the 1793 law, but before Shaw could answer, the women fled the courtroom (with the assistance of the large crowd of abolitionists who filled the courtroom).[48]

Shaw faced the Fugitive Slave Law head-on in 1842. George Latimer was arrested by the Boston police who had a warrant based on a complaint brought by James Gray, a Virginia slaveholder who accused Latimer of being a fugitive and of having stolen Gray's property. Latimer was taken to a Boston jail and held.[49] After an attempt to use habeas corpus to free Latimer was rejected by Shaw (on the grounds that the detention of Latimer conformed to the Fugitive Slave Law), the abolitionists acting on Latimer's behalf sued under Massachusetts's new "personal liberty law," which had been passed in 1837.[50] The statute, in effect, gave fugitive slaves a right to a jury trial on the question of whether they could be held under the Fugitive Slave Law (although it was carefully written so as not to refer specifically to fugitive slaves).[51] An almost

[45] This case, *Commonwealth v Eldridge*, was reported in "several Boston newspapers" during the week of 1 August 1836, according to Levy. See Levy, *Law of the Commonwealth*, n. 21 above, at p. 73.

[46] Ibid., at p. 74.

[47] Shaw asked: "Has the captain of the brig *Chickasaw* a right to convert his vessel into a prison?" and set the women free: *Right and Wrong in Boston in 1836: (Third) Annual Report of the Boston Female Anti-Slavery Society* (1836), p. 51.

[48] Levy, *Law of the Commonwealth*, n. 21 above, at pp. 75–6.

[49] The "Latimer Case" was reported in several Boston newspapers, including the *ad hoc Latimer Journal*, and in (1843) 5 *Law Reporter* 481. See Cover, *Justice Accused*, n. 1 above, p. 169 n*.

[50] Act of 19 April 1837, Massachusetts General Laws 1836–53 Supp., Chap. 221.

[51] Levy, *Law of the Commonwealth*, n. 21 above, at p. 81; Cover, *Justice Accused*, n. 1 above, at p. 164.

identical law had been struck down as unconstitutional by the United States Supreme Court in *Prigg v Pennsylvania* that year.[52] As reported in the press, Shaw stated that, to the extent that the Massachusetts personal liberty law interfered with federal rendition process, it was unconstitutional and void.[53] He said that "he probably felt as much sympathy for [Latimer] as others; but this was a case in which an appeal to natural rights and the paramount law of liberty was not pertinent . . . [i]t was decided by the Constitution of the United States, and by the law of Congress".[54] Shaw refused to free Latimer through the use of Massachusetts law.

Cases like *Latimer* prompted Northern states to pass revised personal liberty laws which did not directly interfere with the federal rendition process, but which prohibited any state officer from assisting in the federal rendition process.[55] By withdrawing state officers from the fugitive slave process, states such as Massachusetts, Pennsylvania, and Ohio made it almost impossible for the 1793 law to be enforced, since there simply were not enough federal judges. The Southern states were furious at what they saw as Northern obstructionism, and the Fugitive Slave Law was amended in 1850 to create a new corps of federal commissioners who would determine whether or not an alleged fugitive should be sent to the state of the slaveholder making the claim.[56] In 1851 Thomas Sims was seized and presented to the new U.S. Commissioner in Boston. James Potter of Georgia, through his agents in Boston, alleged that Sims was a fugitive slave. Sims was kept in the federal courtroom in the Boston Court House which federal marshals barricaded by ringing with chains.[57] Sims's lawyers quickly filed a writ of habeas corpus, which Shaw refused to even entertain on the grounds that if he were to grant the petition, he would simply have to remand Sims to the custody of the U.S. Marshal, since he had no jurisdiction to determine whether Sims was or was not a fugitive.[58] Two days later Shaw relented and agreed to hear argument on whether the Fugitive Slave Law of 1850 was constitutional. He heard argument on Monday, 6 April 1851, and by three o'clock that day he had produced a full opinion.[59]

[52] *Prigg v Pennsylvania*, 16 Peters (U.S.) 539 (1842).

[53] Levy, *Law of the Commonwealth*, n. 21 above, at p. 82.

[54] *The Liberator*, 4 November 1842.

[55] This was permitted by dicta in *Prigg*. See Paul Finkelman, "Story Telling on the Supreme Court: *Prigg v Pennsylvania* and Justice Joseph Story's Judicial Nationalism", (1994) *Supreme Court Review* 247 at 284.

[56] It would be then determined by the slaveholder's state law whether the alleged slave was in fact a slave. See Act of 18 September 1850, 9 Stat. 462 (1850), and Anthony J. Sebok, "Judging the Fugitive Slave Acts", (1991) 100 *Yale Law Journal* 1835 at 1840.

[57] Since the Court House also contained the state courts, the citizens of Boston felt shame and revulsion at the sight of their judges bending beneath the chains to get to their chambers. To the abolitionist Bronson Alcott, the chains were a visible answer to those in Boston who still asked, "What has the North to do with slavery?": quoted in Leonard Levy, "Sims' Case: The Fugitive Slave Law in Boston in 1851", (1950) 35 *Journal of Negro History* 35 at 46.

[58] Levy, *Law of the Commonwealth*, n. 21 above, at p. 95.

[59] *Thomas Sims' Case*, 7 Cush. 285 (Mass. 1851).

Sims had made two arguments: first, that although the Fugitive Slave Clause required Massachusetts to "deliver up" any alleged fugitive slave, it gave no power to Congress to facilitate the process of delivery, and secondly, that if Congress had such a power, only an Article III judge (not a commissioner) could adjudicate claims under it. In his opinion, Shaw stated that there were two questions that he had to answer: first, whether Congress had the authority to pass the Fugitive Slave Law of 1850, and secondly, whether any specific provision of the law violated the Constitution. Levy expressed surprise that Shaw could have written his opinion in just a few hours,[60] but of course, the answer to the first question had been written by Shaw fifteen years earlier, in *Aves*. The first half of *Sims* was simply a restatement of the last section in *Aves*. The only difference is that instead of discussing the Fugitive Slave Clause in dicta, Shaw now discussed it with regard to the holding. The structure of the argument was the same, however.

Shaw began by noting that, before the Constitution was ratified, because of *Somerset*, Northern states were not obliged to extend comity to Southerners in the matter of slavery.[61] If the Constitution had never come about, the only hope a slaveholder would have had for any assistance in the North would have been through a treaty similar to the sort negotiated in international law.[62] If such a treaty had been negotiated, what would it have required of its signatories? Like any treaty, it would have required "the renunciation of some powers of sovereignty" on the question at issue.[63] The issue being slavery, the Northern signatory to such a treaty would have given up the power it had under *Somerset* to extend the normal operation of its criminal and civil laws to *fugitive* slaves, and the Southern state would have gained a guarantee, on behalf of individual slaveholders, that fugitive slaves would be seized in the North.[64]

[60] Levy, *Law of the Commonwealth*, n. 21 above, at p. 98.

[61] *Sims*, n. 59 above, at 296.

[62] Ibid., at 297.

[63] Ibid.

[64] Ibid., at 298, citing *Aves*. It is interesting to ask what, in this hypothetical treaty negotiation, was given up by the South. As Weinberg noted, the threat of "retaliatory comity" is hard to characterise—there were no Pennsylvania slaves escaping into Virginia that Virginia could refuse to return: Weinberg, "Methodological Interventions and the Slavery Cases", n. 7 above, at p.1335. Two possibilities remained. First, a Southern state could have offered, in "exchange" for the Fugitive Slave Clause, to recognise the freedom of a black freed by the operation of Northern law. Within the logic of *Somerset*, entry into a Northern state did not "free" the slave; nonetheless a Southern state could modify its own law to treat voluntary transport to a Northern state as the equivalent of manumission. See *Aves*, n. 23 above, at 369 (discussing *Ex parte Grace*, 2 Hagg. Adm. (G.B.) 94 (1827)) (question of what would should happen if Somerset were to return voluntarily to Louisiana "is a question incidentally raised in the argument, but is one on which we are not called on to give an answer" and *Collins v America, a woman of color* , 9 B. Mon. (Ky.) 565, 571 (1849) (whether to treat black who had lived in free state as free was a matter of Kentucky law)). It is worth noting that in the beginning, Southern states recognised the "liberating" effects of a visit to a Northern state for non-fugitive slaves: see Finkelman, *An Imperfect Union*, n. 3 above, at p. 189 ("in many ways the South was more liberal toward freedom claims than the North" until the mid-1830s). Whether they would have done this but for the guarantee of the return of fugitive slaves is an open question upon which we can only speculate. Secondly, a treaty between a Northern and a Southern state ensuring the recapture of

But, of course, the United States was created, and there was no need for treaties between the states with regard to slavery. Instead, there was the Fugitive Slave Clause in the Constitution. Shaw argued that the Fugitive Slave Clause secured the same ends that his hypothetical treaties would have secured, but with one important difference: the "right" to reclaim a slave was not granted to the citizens of a Southern state "by" a Northern state, it was guaranteed by the federal Constitution. Although Shaw used the conceit of a hypothetical treaty between the states to interpret the meaning of the Fugitive Slave Clause, he never lost sight of the fact that, as a matter of structure, whatever right a slaveholder had was ultimately a *federal* right. Massachusetts had never agreed to modify its criminal and tort law to allow a slaveholder physically to seize alleged slaves with impunity; nor did the federal constitution require Massachusetts to change its laws to allow such seizures. Given Shaw's view that the "right" in slavery was really a complex set of immunities and privileges, not a property right, the idea—urged by Sims's abolitionist lawyers—that the Fugitive Slave Clause was a guarantee that Northern states would devise procedures to determine whether blacks accused of being fugitives were in fact slaves, would have struck Shaw as inconsistent with *Somerset's* premise that, absent positive law recognising slavery, *there was nothing for state fugitive procedures to act upon*. On the other hand, Shaw would have seen nothing strange in the federal congress—in response to *Somerset*—creating procedures for the return of fugitive slaves, since the Congress (regrettably) accepted the institution of slavery in half the country and accordingly passed many "municipal" statutes concerning slavery (including, for example, the ownership of slavery in the District of Columbia).[65] I believe that these considerations explain why Shaw concluded that, like the powers explicitly granted Congress in Article I, section 8, the power to return fugitive slaves was held exclusively by the Congress, even thought the Fugitive Slave Clause is in Article IV and says nothing about Congress.[66]

It was only after the complicated restatement of the theoretical argument he first made in *Aves* that Shaw turned to arguments from history and precedent. He noted that the Second Congress adopted the 1793 Fugitive Slave Law, suggesting that those who had debated and passed the Fugitive Slave Clause thought that it gave Congress the power to regulate the return of fugitive

fugitive slaves would have reduced the incursion by slaveholders into the Northern states. See *Aves* at 371 (such a treaty would reduce "perpetual collisions and border wars") and *Sims* at 296 ("If two states bordered on each other, one a slave state and the other a free state, there would of course be a constant effort of slaves to escape into the free state, and a constant temptation to slave owners to follow and recapture them"). Strictly speaking, the Northern state could have prevented these incursions through the operation of their own laws, but the prophylactic of a promise of rendition would have been more effective than the threat of the arrest of invading slaveholders.

[65] Shaw noted but downplayed the degree to which the Congress itself regulated slavery: "The framers of the constitution could not abrogate slavery, or the qualified rights claimed under it; they took it as they found it, and regulated it to a limited extent": *Sims*, n. 59 above, at 318.

[66] Ibid., at 299.

slaves.[67] He recited the judicial decisions that had upheld the 1793 law: *Wright v Deacon*,[68] *Commonwealth v Griffith*,[69] *Prigg v Pennsylvania*,[70] and *Jones v Van Zandt*.[71] Furthermore, Shaw concluded that in all important respects, the 1850 law stood on the same ground as the 1793 law. Given that the 1793 law did not provide for a jury trial before rendition and allowed non-Article III actors to determine rendition, Shaw did not think that changes made by Congress to the Fugitive Slave Law in 1850 significantly altered its constitutionality.[72] Shaw therefore denied Sims's petition. After further unsuccessful federal appeals, Thomas Sims, under heavy guard, was sent to Savannah from Boston.[73] He was the first alleged fugitive slave ever returned by Massachusetts to the South. Southern newspapers celebrated their victory, while Boston's abolitionists mourned.[74] The abolitionists turned on Shaw with a fury: they accused him of being Pontius Pilate, where Sims the fugitive was Christ.[75]

IV. WHY DID SHAW SEND SIMS SOUTH? COVER'S THEORY

According to Cover, Shaw's decision in *Sims* was consistent with a pattern of "formalistic" decision-making by Northern judges faced with fugitive slave cases.[76] In contrast to his great private law decisions, Shaw's fugitive slave decisions lacked his usual "vigor and surefootedness" and revealed a "very questionable, dogmatic use of precedent".[77] These were but symptoms of Shaw's "retreat to formalism", which was a result of his discomfort at having to adjudicate laws which he found substantively evil.[78] In saying this, Cover did not purport to condemn Shaw. Rather, he sought to demonstrate how his participation in difficult cases like *Sims* caused him to abandon the vigorous and

[67] Ibid., at 300.
[68] 5 Serg. & Rawle 62 (Pa. 1819).
[69] 2 Pick. 81 (Mass. 1823).
[70] 16 Peters (U.S.) 283 (1842).
[71] 13 F. Cas. 1040 (1843).
[72] *Sims*, n. 59 above, at 309.
[73] Upon his arrival, Sims was publicly whipped 39 times. See Levy, "Fugitive Slave Law in Boston", n. 57 above, at 72. During the Civil War, Sims escaped and returned to Boston. He subsequently became an employee in the Department of Justice, working for Attorney General Charles Devens. Ironically, Devens had been the U.S. Marshall in Boston in 1851 and had been responsible for imprisoning Sims and enforcing the decision to send him to Savannah: ibid., at 74.
[74] The *Savannah Republican* noted that it had the "pleasant duty to accord to the authorities and people of Boston great credit for the firm and energetic manner in which they have demeaned themselves": quoted in *Daily National Intelligencer*, 25 April 1851; and *see* Levy, "Fugitive Slave Law in Boston", n. 57 above, at p. 72.
[75] See Cover, *Justice Accused*, n. 1 above, at p. 252. For the first time in Massachusetts, there was serious talk of an elected judiciary, and a state constitutional convention was called in order to consider the option. See (1853) 2 Official Report of the Debates and Proceedings in the Massachussets State Constitutional Convention of 1853, 687–714, 756–832.
[76] Cover, *Justice Accused*, n. 1 above, at p. 251.
[77] Ibid.
[78] Ibid., p. 234.

creative approach he evidenced in other cases in favour of a narrow formalism. Indeed, according to Cover, it was precisely those judges, like Shaw, who "were the most troubled by a result that favored slavery" whose decisions exhibited the greatest degree of formalism.[79] Cover called this the "moral/formal" dilemma.[80]

Cover was very clear about the precise reasons for the moral/formal dilemma. The slave cases forced judges like Shaw to simultaneously think of themselves as "a moral human being, opposed to human slavery . . . and as a faithful judge, applying legal rules" that promoted slavery.[81] The result was "cognitive dissonance" that drove the judge into formalism.[82] This process could be achieved through three different (but mutually reinforcing) strategies. First, a judge like Shaw would "elevate the formal stakes" by choosing "the highest of possible justifications for the principle of formalism relied upon".[83] Cover argued that Shaw did this by beginning his decisions with the premise that without the Fugitive Slave Clause, the Constitution would never have been ratified.[84] Secondly, the judge would "retreat to a mechanistic formalism". This manoeuvre has been closely associated with legal formalism, and was the one which Cover attributed most frequently to Shaw.[85] According to Cover, Shaw's legal reasoning was "mechanistic" because he wrote as if the legal result in *Sims* was compelled by precedent or legislative intent, and ignored the fact that "like all legal issues of complexity, [it] was amenable to a broad range of solutions with a concomitant broad area for potential introduction of morality".[86] The final strategy that a judge might employ was the "ascription of responsibility elsewhere".[87] By focusing on the legal principle of separation of powers, Shaw was able to "externaliz[e] responsibility for unwanted consequences" by blaming Congress or the framers of the Constitution.[88] In other words, Shaw-the-judge had to construct a very narrow role for adjudication in order to prevent total cognitive dissonance with Shaw-the-abolitionist.[89] Cover's point was that by the time Shaw was finished resolving his cognitive dissonance, he had produced

[79] Cover, *Justice Accused*, n. 1 above, at 229.
[80] Ibid., pp. 198–9.
[81] Ibid., p. 228.
[82] Ibid., p. 229.
[83] Ibid., p. 199.
[84] Ibid., pp. 265–6.
[85] Ibid., p. 234.
[86] Ibid., p. 233.
[87] Ibid., p. 236.
[88] Ibid., p. 236.
[89] Of all the judges Cover discussed, Shaw seemed to affect him most deeply. He begins his book by trying to connect Shaw to Melville's literary creation Captain Vere in *Billy Budd* (Shaw was Melville's father-in-law): "Melville, in his last artistic act, may have been paying his respect to the soul of his departed father-in-law to whom he dedicated in more conventional fashion his first novel [*Typee*]. If so, Melville, as least, saw the soul of Shaw as intertwined with the complex interaction of crucifixion forgiveness and salvation. Such theological symbolism would be unlikely without the more earthly concomitant of guilt, remorse, and justification": Cover, *Justice Accused*, n. 1 above, at p. 252.

a set of decisions whose reasoning bore almost no relation to his best decisions. Instead, Shaw's slave decisions which relied upon implausible and extreme claims for the determinacy of legal reasoning.

Cover's analysis of Shaw's response to the fugitive slave cases in the North has been accepted by many modern scholars as thorough and persuasive. Nonetheless, there is something curious about Cover's "discovery" of a deep vein of extreme formalism in Shaw. As Cover himself noted, scholars since Karl Llewellyn had characterised Shaw and other early nineteenth century judges as *anti*-formalists.[90] Cover agreed with Morton Horwitz's thesis that the antebellum period was marked by an "instrumental" approach to adjudication, which Llewellyn had called the "grand style" of legal reasoning.[91] According to the view propounded by Llewellyn and Horwitz, instrumentalism was a pragmatic policy-oriented approach to adjudication in which judges were driven by "considerations of policy or 'convenience,' [and] the functional needs of society".[92] Instrumentalism, in short, is the opposite of formalism, and Shaw was seen as one of the pioneers of instrumentalism in private law.

Although Cover did not think that Llewellyn and Horwitz were wrong about the distinction between instrumentalism and formalism, he was concerned that their model of formalism had been applied too crudely: "[judicial] appeals to formalism may be not only the product of an 'age' . . . Thus, in slavery . . . the 1840's and 1850's were not a golden age of free-wheeling policy jurisprudence, but an age of the retreat to formalism".[93] I do not think that Cover was contradicting Horwitz's thesis when he tried to prove that Shaw was a formalist in the slave cases. Horwitz and Cover understood the term "formalism" in the same way, but Cover sought to establish that, with respect to the agonizing issues of the fugitive slave cases, Shaw's judicial practice did not follow his private law decisions.

V. WHY DID SHAW SEND SIMS SOUTH? NELSON'S THEORY

By contrast, Nelson—who was self-consciously working within the Horwitz/Llewellyn model—offered an account of Shaw's slave law decisions directly contrary to Cover's. Indeed, he concluded that that Shaw and other Northern judges had been unable to protect fugitive slaves not because they were formalists but because they were not formalist enough.

[90] Cover, *Justice Accused*, n. 1 above, at p. 200; see also Karl Llewellyn, *The Common Law Tradition: Deciding Appeals* (Boston: Little, Brown, 1960).

[91] Cover, *Justice Accused*, n. 1 above, at p. 200 (citing Morton J. Horwitz, "The Emergence of the Instrumental Conception of American Law, 1780–1820", (1971) *5 Perspectives in American History* 285 at 287.

[92] Robert W. Gordon, "The Case for (and Against) Harvard", (1995) 93 *Michigan Law Review* 1231 at 1252–3.

[93] Cover, *Justice Accused*, n. 1 above, at p. 200.

To Nelson, the fugitive slave cases also posed a moral/formal dilemma. But the dilemma he described was the opposite of Cover's. According to Nelson, natural law arguments on behalf of slaves like Sims (which Cover thought constituted the "moral" side of the dilemma) were "formalist" in that they posited fixed and binding rules which would trump policy arguments. By contrast, the argument for rendition, which Cover took to consist of a formalist reliance on precedent, was conceived by Nelson as a normative—*utilitarian*—argument for "the preservation of the Union".[94] Thus, according to Nelson, judges like Shaw treated the slave cases with the same instrumentalist creativity that they applied to common law cases involving private law: they focused on "the promotion of economic growth by deciding specific cases in a 'manner most conducive to the general prosperity of commerce' ".[95] Shaw and the other Northern judges "rested their case upon instrumentalist arguments about what was politically wise and economically expedient, whereas opponents of slavery made essentially moralistic arguments about the law of God and the rights of man".[96] Far from being the foundation of Shaw's slavery decisions, Nelson concluded that formalism instead arose in reaction to the instrumental judicial style of *Latimer* and *Sims*:

> "As a result of the association of instrumentalism with proslavery forces before the war and the political defeat of those forces during the 1860's, instrumentalism became discredited as a style of judicial reasoning, thereby creating a void that had to be filled. That void, as will appear, was ultimately filled by American formalism, which . . . enabled judges to avoid engaging in the sort of utilitarian and political reasoning that had been commonplace to instrumentalism".[97]

For Cover, Shaw was a formalist because he mechanistically applied positive federal law and refused to permit his own political and moral judgements to play a role in his adjudication of the fugitive slave cases. For Nelson, Shaw was an instrumentalist because he permitted his political and moral judgements to play a role in his adjudication of the fugitive slave cases. According to Cover, formalism helped Shaw rationalise his refusal to exercise his will, while according to Nelson, formalism would have been a brake on Shaw's wilfulness. Cover

[94] Nelson, "Impact", n. 2 above, at 540.

[95] Ibid. at 514 (quoting *Thurston v Koch*, 23 F. Cas. 1183, 1186 (No. 14,016) (C.C.D. Pa. 1805)). For Nelson, early nineteenth century instrumentalism was perfectly symbolised by *Charles River Bridge v Warren Bridge*, 36. 11 Pet. (U.S.) 420 (1837), aff'g 7 Pick. 344 (Mass. 1829), in which, like "innumerable contemporary cases, the judges chose to modify rules to promote development rather than to have stable and predictable rules": Nelson, "Impact", n. 2 above, at 519. *Charles River Bridge* also played an important role in Horwitz's discussion of instrumentalism: he used it to prove how, under instrumental reasoning, "conventional notions of property rights began to give way under the pressures of economic development": Morton J. Horwitz, *The Transformation of American Law* 1780–1860 (Cambridge: Harvard University Press, 1977), p. 132 Clearly, Nelson agreed with the Horwitz/Llewellyn model both as to its definitions and its dates, as did Cover. Nelson and Cover simply had diametrically opposed views as to whether the slave law instantiated formalism or instrumentalism.

[96] Nelson, "Impact", n. 2 above, at 544.

[97] Ibid., at 548.

thought that the abolitionist, if he had a jurisprudence, would reject formalism, while Nelson thought that he would embrace formalism.

VI. READING *SIMS* AND *AVES* TOGETHER

The conflict between Cover and Nelson cannot be resolved in their terms because both began and ended their study of Shaw's slave law jurisprudence with Shaw's fugitive slave cases.[98] This may have seemed to them like a reasonable place to have begun, but it is not, since Shaw viewed the fugitive slave cases as integrally connected with his earlier decisions in the slave transit cases. As I have shown, the main argument—even the language—of *Sims's Case* had been set out by Shaw in *Aves* in 1836. Whatever Cover or Nelson found in *Sims* should be in *Aves* as well. But, as I will argue below, it is very difficult to characterise *Aves's* treatment of the Fugitive Slave Clause as either formalist or instrumentalist.

Although Cover never passed judgement on Shaw's legal reasoning in *Aves*, the fact that its treatment of the Fugitive Slave Clause was identical to *Sims* suggests that he should have thought it a formalist decision, notwithstanding its pro-liberty result. This gives rise at once to a question about Cover's causal explanation: if Shaw's formalism in *Sims* was a result of "dissonance" between his hostility to slavery and the palpable demands of the pro-slavery Constitution, what was the cause of his "formalism" in *Aves*, where the dissonance would have been much less pronounced? The circumstances that, according to Cover, created the psychological pressure for Shaw's "retreat to formalism" were not nearly as strong in 1836 compared to 1851. In *Aves* Shaw was not sending a man back to slavery—he was giving a man his freedom. There was no practical conflict between Shaw's view of himself as a moral man and his view of himself as a faithful judge. Furthermore, while there was conflict over slavery and over the fate of fugitive slaves, the degree of that conflict was much less in 1836 than in 1851. The legal and political landscape was far less polarised in 1836.[99] There had been only a handful of decisions concerning either the constitutionality or interpretation of the Fugitive Slave Law of 1793, and it is not clear just how salient the issue was in the North.[100] In fact, until the mid-1830s

[98] Cover referred to *Aves* once in *Justice Accused* (in a discussion of the reception of *Somerset's Case* in the USA). See Cover, *Justice Accused*, n. 1 above, at p. 94. This single reference does not mention Shaw's reliance on the Fugitive Slave Clause to interpret *Somerset*. *Aves* was subsequently never mentioned in Cover's extensive discussions of Shaw's formalism, his views on the Fugitive Slave Clause, and his decision in *Sims*. Nelson did not discuss *Aves* in "Impact", n. 2 above.

[99] See J. Smith, "The Federal Courts and the Black Man in America, 1800–1883" (unpublished Ph.D. dissertation, University of North Carolina, Chapel Hill, 1977), at 164–5 (national fugitive slave policy "stirred relatively little controversy among public until 1830s"); Nelson, "Impact", n. 2 above, at 533 (Northern elites did not come to see Southern domestic slave laws and practices as a threat to the integrity of the domestic affairs of Northern states until after 1836).

[100] See, e.g., *Butler v Hopper*, 4 F. Cas. 904 (1806) (Fugitive Slave Law did not apply to slaveholder who voluntarily brought slave into Pennsylvania); *In re Susan*, 23 F. Cas. 444 (1818) (alleged

the South was recognising the effects of Northern emancipation law: slaves who were voluntarily brought North often sued for their freedom in Southern States and won.[101] As Cover noted, the dramatic change that occurred in the abolitionist strategy towards slavery began in the mid-1830s but did not really gain momentum until the 1840s and 1850s.[102] There is no reason to assume, therefore, that Shaw suffered the same kind of intense cognitive dissonance in *Aves* that Cover claimed he must have suffered in *Sims*.

It is also worth asking whether any of the three "formalist" strategies that are supposed to have produced *Sims* are apparent in *Aves,* and if not, why not. Taking them in order, we see that only one of them is obviously part of the simplest or most satisfying interpretation of the decision. First, Cover argued that in *Sims* Shaw "elevated the formal stakes" by arguing that without the Fugitive Slave Clause the Constitution would not have been ratified. In *Aves* Shaw did not argue that *without* the Fugitive Slave Clause the Constitution would not have existed, but rather, he used a counterfactual to determine what the Fugitive Slave Clause meant, given the fact that the Constitution *was* ratified. In fact, Shaw's technique of trying to deduce from the principles of political science what the authors of the Constitution "must" have meant is not very different from the sort of arguments employed by Hamilton in the *Federalist Papers* and Marshall in the early years of the Supreme Court.[103] Shaw did not think that the interpretation should be based exclusively or even mostly on the subjective intentions of the Fugitive Slave Clause's actual authors. Instead, like Marshall, Shaw took the actions of the framers as a starting point of analysis, and attributed to them only those intentions that could be reasonably or rationally attributed to a corporate body trying to achieve the general aims of the Constitution

fugitive returned from Indiana without reported opinion); *Wright v Deacon*, 5 Serg. & Rawle 62 (Pa. 1819) (upholding Fugitive Slave Law); *Commonwealth v Griffin*, 2 Pick 11 (Mass. 1823) (same); *Jack v Martin*, 14 Wend. 507 (N.Y. 1835) (denying that Fugitive Slave Law was constitutional but ordering return of alleged slave under authority of Fugitive Slave Clause).

[101] See Finkelman, *An Imperfect Union*, n. 3 above, at p. 181 ("A surprising number of slaves were freed by courts in slave states, especially in the period before 1840"); *Rankin v Lydia*, 2 A.K. Marsh. 467, 470 (Ky. 1820) (if by "positive provisions of our code, we can and must hold our slaves in the one case, and statutory provisions equally decide against that right in [another state], and liberate the slave, he must . . . be declared free") and see n. 7, above.

[102] See Cover, *Justice Accused*, n. 1 above, at pp. 160–2. *Prigg* (1842) may have been a powerful catalyst. As Finkelman noted, after *Prigg* there was an explosion in Northern states of "personal liberty laws" which prohibited state officials and/or provided for certain state procedural protections for alleged fugitives: Finkelman, "Story Telling on the Supreme Court", n. 55 above, at 252–3, 290.

[103] See G. Edward White, *The Marshall Court and Cultural Change*, 1815–1835 (New York: Oxford University Press, 1991), pp. 124–5 (on the debt of Marshall's theory of "coterminous powers" to the *Federalist Papers*). One of Marshall's most important decisions, *Cohens v Virginia*, 6 Wheat. (U.S.) 264 (1821), exemplified the style of argument developed in the *Federalist Papers*: in place of technical legal analysis, "the only case reference in the seventy-three-page opinion is to *Marbury*. In support of the decision, Marshall cites only the *Federalist*, the Judiciary Act, the Constitutional Convention, and a panoply of appeals to reasonableness, arguments from the 'nature of government' and the 'nature of the Constitution', from the 'general spirit of the instrument' ": Shannon Stimson, *The American Revolution in the Law* (London: Macmillan Press, 1990), p. 140 (quoting Cohens).

as expressed through both its structure and its language.[104] Thus, given the Constitution's innovative federal structure, and the unique problems created by *Somerset*, Shaw's belief that the Fugitive Slave Clause gave exclusive power to the federal government was based on his construal of the text, not on any privileged historical insight.[105]

Secondly, Cover argued that in *Sims* Shaw relied on the idea that precedent or legislative intent compelled a rendition. In *Aves* Shaw confronted and rejected the widespread assumption that *Somerset* was not the correct rule of law and that comity should be extended to Southern slaveholders in transit in the North. Shaw's adoption of *Somerset* in *Aves* could hardly be called mechanistic, and it is hard to see how the affirmation of that decision became mechanistic fifteen years later. As in his treatment of tort and property law, Shaw's main concern was to reconcile conflicting strands of doctrine by subsuming them under the most defensible general principle of law that explained the most number of cases. It was under this model of legal reasoning that Shaw invoked the Constitution as evidence in favour of his radical correction of Massachusetts's conflict of laws doctrine. Rather than being mechanistic, Shaw's method was synthetic and interpretive, and looked very similar to his best work in private law.[106]

Thirdly, Cover argued that in *Sims* Shaw "externalized" responsibility by blaming Congress or the framers of the Constitution for the decision to refuse to extend aid to Sims. I think *Aves* cannot be used to rebut this allegation; in fact here Cover was clearly right. Even in *Aves*, where Shaw was in the happy position of using *Somerset* to free a slave, Shaw wrote as if he was only obeying the rules of the common law, and that he would not have been doing so if there were any valid municipal law (like the Fugitive Slave Law of 1793) telling him otherwise. In two ways, Shaw was placing responsibility "elsewhere": first on the common law (*Somerset* was "the law") and second on the Massachusetts legislature and the Constitutional Congress (slavery could be created by municipal

[104] Shaw's method was not unlike the "structural" approach championed by Chief Justice Marshall in *Marbury v Madison*, 1 Cranch (U.S.) 137 (1803) and *McCulloch v Maryland*, 4 Wheat. (U.S.) 316 (1819). See Harry H. Wellington, *Interpreting the Constitution* (New Haven: Yale University Press, 1990), pp. 21–4 (comparing Marshall's structuralism with that developed by Charles Black); Philip Bobbitt, *Constitutional Fate* (New York: Oxford University Press, 1982) , pp. 78–9 (same); but see Robert Nagel, *Constitutional Cultures* (Berkeley: University of California Press, 1989), p. 133 (Marshall's method in *McCulloch* was "a sophisticated literalistic argument that largely obliterated the original text").

[105] As James Boyd White noted, this is not unlike how a "modern constitutional lawyer" would approach the Fugitive Slave Clause. See James Boyd White, "Constructing a Constitution: 'Original Intention' in the Slave Cases", (1987) 47 *Maryland Law Review* 239 at 245 (comparing Story's interpretation of the Fugitive Slave Clause with a "modern" approach to federal powers).

[106] The idea that Shaw approached the Fugitive Slave Clause "mechanistically" is difficult to square with one of the first things that Shaw says at the beginning of *Sims*: "The constitution of the United States is not to be expounded as if it were now opened for the first time, and with sole regard to the words and figures in which is expressed; its history is too deeply interwoven with our whole social system, to be disregarded, when we are called upon to ascertain its full meaning and effect": *Sims*, n. 59 above, at 295.

law). In *Aves*, Shaw was willing to externalise responsibility for *wanted* consequences as well as unwanted consequences. Whether the fact that Shaw clearly embraced the idea that the judge was not the source of law made him a formalist (as opposed to merely a positivist) will be discussed below.

Now let us examine the claim that Shaw was an instrumentalist. As above, a critic like Nelson can make his case only by ignoring the remarkable continuity in Shaw's thinking from 1836 to 1851. Shaw did not change his arguments to fit the needs of the times: he predicted in *Aves* how he would treat federal fugitive slave laws, and he followed through on his prediction even when the politics of the decision were fraught with peril.[107] If one feature of instrumentalism is a penchant to treat legal rules as fluid in order to conceal judicial activism, then it is hard to see how Shaw was an instrumentalist. One might argue that Shaw's decision to reject comity in slave transit cases in favour of the rule in *Somerset* was "instrumentalist" in the same way that *Brown v Kendall* or *Charles River Bridge* were instrumentalist—that Shaw, while careful not *frequently* to change the law, made a few strategic changes that dramatically changed its course. Thus, by imposing *Somerset* onto Massachusetts, Shaw forced slave law to change according to his vision of the good.

It is not clear why a judge who reverses precedent in a major case is obviously an instrumentalist. We might think, on the contrary, that such a judge was in fact bringing the law back to where it always should have been.[108] Notwithstanding this possibility, Nelson suggested that the judges of the first half of the nineteenth century were instrumentalist because of the *reasons* for their dramatic reversals of precedent. In property and tort law, instrumental judges reversed entrenched precedents in order to promote economic growth. The argument for viewing *Aves* and *Sims* as instrumentalist decisions is that they served policy goals similar to those of economic growth—not wealth, *per se* but "national unity".[109] According to Nelson, the pro-industrialisation decisions of the 1820s and 1830s were simply one aspect of a broader category of "policy" which, in the slave decisions, manifested itself as a concern for the preservation of the nation at any cost. Thus, in Nelson's mind, there was no difference between Shaw's decision in *Sims* and Taney's decision in *Scott v Sandford* because both were motivated by a judicial obsession with national unity.[110]

[107] See Levy, *Law of the Commonwealth*, n. 21 above, at p. 73.

[108] "When Chief Justice Shaw ruled in *Commonwealth v Aves* that a slave voluntarily brought into Massachusetts was free, he was not creating new law, he was returning the law of Massachusetts to its earlier philosophical position. Indeed, Shaw was surprised to discover that his decision was novel or precedent setting": Finkelman, *An Imperfect Union*, n. 3 above, at p. 341.

[109] Nelson, "Impact", n. 2 above, at 541. Levy argued that, in the minds of the abolitionists, at least, the goal of national unity was never very far from economic self-interest: "By mid-century there was no longer any concealment by a great majority of Boston's 'best people' of their warmness towards Southern interests": Levy, "Fugitive Slave Law in Boston", n. 57 above, at 40; and *see* Levy, *Law of the Commonwealth*, n. 21 above, at pp. 102–3 (on "hunkerdom", the merchant class of Boston that rejoiced when Sims was sent South).

[110] 60 U.S. (19 How.) 393 (1857) and see Nelson, "Impact", n. 2 above, at 544. Ironically, Cover's first formalist strategy, the "elevation of the formal stakes" is indistinguishable from Nelson's

An initial problem with Nelson's theory is that his hypothesis that Shaw was concerned with "national unity" does not in itself prove very much. Obviously, Shaw was concerned in both *Aves* and *Sims* with proving that the federal government had exclusive power over fugitive slaves. And it should be equally obvious that Shaw, in defending the Fugitive Slave Laws, was also expanding federal power. This would explain why Story approved of *Aves* and Shaw, in turn, approved of Story's decision in *Prigg*.[111] Both Shaw and Story were, to a great extent, followers of Marshall and they tried to carry forward the federalist project in an era that was increasingly hostile to federalism. But we must not equate *federalism* with "national unity". By "national unity" Nelson meant (ultimately) national *survival*. Nelson thought *Dred Scott* symbolised the judiciary's desire to preserve national unity, and argued that, no less than *Dred Scott*, *Sims* was an instrumentalist decision.[112] Nelson saw no difference between Chief Justice Taney's motives, which were to please the South at all costs, and Shaw's motives in *Sims*. In his zeal to prove that every pro-slavery decision by Northern judges was cut from the same instrumentalist cloth, Nelson ignored the simple and critical fact that while Shaw's decisions were decidedly federalist, Taney's decision in *Dred Scott* was unabashedly anti-federalist.[113]

In fact, except for the fact that *Sims* and *Dred Scott* both involved slavery, the holdings in each case have nothing to do with each other. *Sims* dealt with an alleged fugitive slave. *Dred Scott* concerned a slave who had visited a free state with his master. If *Dred Scott* is to be compared with any of Shaw's decisions, it should be *Aves*. But even the most casual comparison of the two cases reveals that Shaw's reasoning in *Aves* was the opposite of Taney's reasoning in *Dred Scott*. Shaw upheld the freedom of Med in *Aves* based on *Somerset*. Taney rejected Scott's claim to freedom partly (although not mainly) because he rejected the idea that slaves in transit through the territories could gain their

characterisation of instrumentalism in the fugitive slave cases. According to either theory, the judge in question (for example, Shaw) refers to a catastrophic alternative as the ground of his decision. For Cover such a move is formalist because it allows the judge to (non-heroically) displace responsibility on fate or history; while for Nelson such a move is instrumentalist because it allows the judge to (heroically) avert a result to which the nation was condemned by fate or history.

[111] Story's decision in *Prigg* reflected his commitment to federalism. See James Boyd White, "Constructing a Constitution", n. 105 above, at 248 (*Swift v* Tyson, 41 U.S. (16 Pet.) 1 (1842) was Story's "version of judicial nationalism, built on Marshall's but going beyond it") and Christopher L. M. Eisgruber, "Justice Story, Slavery, and the Natural Law Foundations of American Constitutionalism", (1988) 55 *University of Chicago Law Review* 273 at 326 (comparing Story's commitment to federalism with Marshall's). It must be noted that we do not know how much of *Prigg* Shaw endorsed. He probably would not have agreed with Story's conclusion that the Fugitive Slave Clause gave individual slaveholders the right to recapture their slaves in a Northern state even without the benefit of federal law. See Finkelman, "Story Telling on the Supreme Court", n. 55 above, at 252.

[112] See Nelson, "Impact", n. 2 above, at 544 (on *Dred Scott*) and at 540 (on *Sims*).

[113] In addition to denying that a black could be a citizen of the USA, in *Dred Scott* the Supreme Court struck down the Missouri Compromise of 1820, thereby taking away from Congress the power to legislate for the territories: *Dred Scott v Sandford*, 60 US (19 How.) 393 (1857) at 446–7. Taney removed from Congress an important power that it had exercised since 1787: see Don E. Fehrenbacher, *The Dred Scott Case: Its Significance in American Law and Politics* (New York: Oxford University Press, 1978), p. 384.

freedom through *Somerset*. Furthermore, while not prohibiting the Northern states from applying *Somerset*, Taney weakened its effect by barring blacks from suing in the federal courts.[114]

While we have no direct evidence of what Shaw thought of *Dred Scott*, we do know that one of its two dissenters, Benjamin Curtis, relied on *Aves* for his attack on Taney's decision. Ironically, it was Curtis, who had twenty years earlier represented the slaveholder in *Aves*, and had argued that Massachusetts should give effect to Louisiana's positive law of slavery just as it would give effect to a foreign state's positive law of real property. Shaw rejected that argument, as we saw above, on the grounds that since slavery was a complex package of criminal privileges and tort immunities, it could only exist if Massachusetts chose to modify its law to create it. In *Dred Scott*, Curtis adopted Shaw's argument to critique Taney's: he pointed out that a slave becomes free (except if a fugitive) in a Northern state not because of the operation of any municipal law, but because of the *lack* of any municipal law under which any other man might imprison or coerce him.[115] Curtis used this fact to then buttress his argument for the constitutionality of Congress's power to regulate slavery in the territories through municipal legislation. He argued that it would have made no sense for the framers of the Constitution, who knew about *Somerset*, to have "conferred the right on every citizen to become a resident on the territory of the United States with his slaves, and there to hold them as such, but [have] neither made nor provided for any municipal regulations which are essential to the existence of slavery".[116] It was more "rational", Curtis argued, to assume that Congress retained for itself the power to make such municipal laws, and, by extension, retained the power to choose to make no law at all, or, as it did, to positively prohibit slavery.[117] Obviously, Curtis's dissent contained all the elements of Shaw's theory of the Fugitive Slave Clause set out in *Aves* and *Sims*: that slavery could only exist as a matter of positive municipal law; that, per *Somerset*, slavery could not extend outside of the South without positive law; that the federal constitution must be read as a document designed to rationally respond to *Somerset*; and that the most rational method of dealing with *Somerset* in a federal system would have been to give the federal government exclusive power to legislate on slavery in those parts of the USA in which there was no domestic municipal law *creating* the institution of slavery.

It is difficult to see, therefore, in what way Shaw's approach to the problem of fugitive slaves reflected the view that law "was essentially mutable and tran-

[114] See Finkelman, *An Imperfect Union*, n. 3 above, at p. 281.

[115] *Dred*, n. 133 above, at 591 (citing *Aves*). Levy suggested that Curtis "reversed" his 1836 position and "followed Shaw's decision": Levy, *Law of the Commonwealth*, n. 21 above, at p. 64. This seems clearly right.

[116] *Dred*, ibid., at 625.

[117] *Dred*, ibid., at 625–6 and see Alfred Brophy, "Let Us Go Back and Stand Upon the Constitution: Federal-State Relations in *Scott v Sandford*", (1990) 90 *Columbia Law Review* 192 at 210 (noting Curtis's federalist approach to the question of Congressional control of slavery in the territories).

sitory in nature".[118] Neither of Nelson's two features of instrumentalism played a prominent role in Shaw's legal reasoning. Shaw did not uphold the constitutionality of the Fugitive Slave Laws of either 1793 or 1850 because of his interest in achieving a specific social policy in spite of what the "law" required. Shaw first developed his theory of the meaning of the Fugitive Slave Clause in the context of a case in which he freed a slave and angered the South; he further developed the theory in a series of cases in which he returned an alleged slave to the South and angered the abolitionists of his own state; and if the reasoning in Justice Curtis's dissent can be taken as a fair reflection of *Aves* (as I think it can), Shaw would have recognised that the logic of his theory would have led him to free a slave and anger the South in *Dred Scott*.[119] Secondly, Shaw did not claim at any point that he believed that his decisions in either 1836 or 1851 would promote "national unity". Not only, as we saw above, would it be anachronistic to claim that it was apparent in 1836 that the fugitive slave question would become critical to the preservation of the nation; even if a prescient observer could have made such a prediction, the language of Shaw's decisions simply do not support the conclusion that he felt torn between the law and national unity. To put it bluntly, Shaw did not see the "purpose" of the Fugitive Slave Act as promoting national unity; its purpose was to fulfil a responsibility that, because of *Somerset*, had been given to the federal government by the Constitution. Once we see that Shaw viewed the Fugitive Slave Clause as a federalist response to *Somerset*, we can see why Shaw argued in both *Aves* and *Sims* that there was no conflict between the purpose of the Fugitive Slave Laws and the Constitution. Contrary to Nelson's assertion, Shaw did not construct his decision in *Sims* as a choice of policy *over* law.

VII. SHAW, FUGITIVE SLAVES, AND LEGAL POSITIVISM

It is clear, therefore, that Shaw's reasoning about the problem of fugitive slaves was neither formalist nor instrumentalist. How can we characterise it then? To understand Shaw's reasoning, and its proper classification, we should recall the two aspects of the critiques above that survived inspection. First, Cover was right to accuse Shaw of "externalizing" the responsibility for the results of the Fugitive Slave Laws. In *Latimer* Shaw stated that he "probably felt as much sympathy for [Latimer] as others" but he would enforce the law notwithstanding Latimer's appeals to "natural rights and the paramount law of liberty".[120] Secondly, Nelson was right to observe that, in general, Shaw was willing to

[118] Nelson, "Impact", n. 2 above, at 520.

[119] As Aviam Soifer noted, Shaw's treatment of a slave brought voluntarily by her master into Boston in 1857 suggests that he "was defying, or at the very least ignoring, the holding in *Dred Scott*": Aviam Soifer, "Status, Contract, and Promises Unkept", (1987) 96 *Yale Law Journal* 1916 at 1923 (citing *Betty's Case*, 20 Monthly L. Rep. 455 (Mass. 1857)).

[120] *The Liberator*, 4 November 1842 (reporting Shaw's statements).

reverse deeply entrenched precedents and to exercise judicial power to change the course of the law. His decision to insist on the rule in *Somerset* reversed, as Finkelman noted, an assumption about the extension of comity to slaveholders in transit that was fifty years old.[121] Shaw's belief that the content of the law concerning fugitive slaves could be found in a set of legal sources and his insistence on applying *Somerset* even though it had been ignored for fifty years implies that Shaw viewed law as a set of rules whose authority depended very much on a specific kind of non-moral criteria. Furthermore, Shaw's insistence on applying the law even though it was immoral implies that Shaw believed that the morality and the existence of a law were separate matters, and could conflict. The idea that legal validity depends on a specific set of sources; that law is a system of rules; and that it has no necessary connection between law and morality is, of course, the basic content of legal positivism.

Classical legal positivists like Bentham and Austin first articulated the three basic tenets of positivism.[122] The first tenet was the separability thesis: that there is no necessary connection between law and morality.[123] The second tenet was the "command theory of law": that law was an expression of human will. Bentham believed that a law must be reducible to a command that one person might give another.[124] In contrast, Austin was not as sure as Bentham that a law had to be reducible to a verbal form. He recognised that some intelligible commands could be merely expressive.[125] The third tenet was the "sources thesis": that every valid legal norm was promulgated by the legal system's sovereign, and that the norm's authority could be traced to that sovereign. According to Bentham, "the authenticity of a law is a question exterior to, and independent of, that of its content", and one therefore had to know by whom and in what manner a norm was promulgated in order to determine its status as

[121] Finkelman, *An Imperfect Union*, n. 3 above, at p. 341.

[122] Classical positivism developed in reaction to classical common law theory. The following represents a general description of the principles proposed by Bentham and Austin. This list is not likely to be precise or exhaustive, but it should suffice for purposes of this discussion. It is compiled from H.L.A. Hart's discussion of classical positivism as well as other sources. See H.L.A. Hart, "Positivism and the Separation of Law and Morals", (1958) 71 *Harvard Law Review* 593 at 594–606; H.L.A. Hart, *The Concept of Law*, 2nd edn. (New York: Oxford University Press, 1994), pp. 18–78; Jules L. Coleman and Jeffrie G. Murphy, *Philosophy of Law*, 2nd edn. (Boulder, Colo.: Westview Press, 1990), pp. 19–33. It should be pointed out that all these works discuss positivism in terms of their reaction, to some extent, to the theories of John Austin. See John Austin, *The Province of Jurisprudence Determined* (London: Weidenfeld & Nicolson, 1955); John Austin, *Lectures on Jurisprudence* (London: J. Murray, 1861).

[123] Bentham attacked Blackstone's use of natural law to explain the authority of common law. The appeal to natural law was not only an appeal to an unprovable "chimera", but it also allowed each law-applier to inject his own morality into the law: Gerald J. Postema, *Bentham and the Common Law Tradition* (New York: Oxford University Press, 1986), p. 269 (quoting Bentham Manuscripts in the University College London Library).

[124] Jeremy Bentham, *The Works of Jeremy Bentham* (Bowring (ed.), Edinburgh: William Tait, 1838–43), vol. 8, p. 94.

[125] "A command is distinguished from other significations of desire, not by the style in which the desire is signified, but by the power and purpose of the party commanding to inflict an evil or pain in case the desire be disregarded": Austin, *Province*, n. 122 above, at p. 14.

law.[126] Austin built upon the sources thesis as set out by Bentham and refined the definition of the sovereign.[127] The key point for Austin, as for Bentham, was to discover the unique *social* source of legal norms in a given legal system.

Classical legal positivism was a powerful theory but it got many things wrong. The most important mistake it made was in the command theory of law. As H.L.A. Hart demonstrated, Austin's idea that law had to be an expression of a human "will" introduced many confusions into legal positivism and fatally weakened it as a theory of law for modern societies like England or the USA. Instead of defining law as a command, Hart suggested that the positivist view law as a "social rule". A social rule is a "convergent social practice and a shared critical or reflective attitude towards that practice". [128] Hart argued that if modern legal systems are built out of "social" rules, not commands, then legal positivism must explain how different rules are created, changed, and identified. Hart built on the insight that law both limits and expands liberty, and argued that rules that limit liberty are "primary" rules, while rules that confer powers are "secondary" rules.[129] He then distinguished between three types of secondary rules. There are rules that confer the power to change primary rules. Hart included in this category the rules of legislation and the powers of private law-making ("the making of wills, contracts, transfers of property").[130] There are rules that confer the power to adjudicate primary rules. These rules determine whether "a primary rule has been broken".[131] But the most important type of secondary rule is the "rule of recognition". This rule sets out the conditions that must be satisfied by a norm if it is to count as a primary rule.[132] In addition to the social rule thesis, which replaced the command theory, Hart's modern positivism still accepted, and needed, the separability thesis and the sources thesis. Social rules might still violate morality, but every social rule still had ultimately some source (as set out by the rule of recognition) under which it was validated.

That Shaw embraced the separability thesis is obvious, and it is something upon which both Cover and Nelson agreed. It is more difficult to determine whether Shaw embraced the sources thesis and the command theory. I will argue that Shaw embraced a nineteenth century version of the sources thesis,

[126] Postema, *Bentham and the Common Law Tradition*, n. 123 above, at p 313 (quoting Bentham Manuscripts in the University College London Library).

[127] According to Austin, the sovereign was identifiable by two characteristics: habitual obedience from the bulk of the population, and habitual non-compliance with the commands of any other human superior. See Austin, *Province*, n. 122 above, at pp. 193–4. For Austin, who was most concerned with modern constitutional democracies, the source of law lay with that body of people that has ultimate authority to alter the state's constitution (the population at large): ibid., pp. 250–1.

[128] Hart, *Concept of Law*, n. 122 above, at pp. 265–6.

[129] Ibid., p. 94.

[130] Ibid., p. 96.

[131] Ibid., p. 97.

[132] Ibid., p. 94–5. As Coleman and Leiter note, unlike the other two secondary rules, the rule of recognition is not a power-conferring rule. See Jules L. Coleman and Brian Leiter, "Legal Positivism", in Dennis Patterson (ed.), *A Companion to Philosophy of Law and Legal Theory* (Cambridge, Mass.: Blackwell, 1996), p. 245.

and that he rejected the command theory in favour of an approach that looks a little like the social rule thesis.

As Cover noted, the "formal assumptions" of the judiciary in early nineteenth century America included the view that both the Constitution and the common law constrained judicial wilfulness.[133] Judges like Story and Shaw were "constitutional positivists" in that they believed that a constitution, being a superior law compared to legislation, would provide judges with a *legal* basis upon which to strike down legislation that conflicted with the constitution.[134] Judges would be prevented from acting wilfully since they would be required to find the *source* of authority to exercise judicial review in the constitution itself.[135] Judges like Story and Shaw developed a very specific view of common law which was developed, to a great extent, in reaction to the codification movement that had begun to gather support during the 1820s and 1830s.[136] The codifiers distrusted common law because it allowed unelected judges to exercise too much power.[137] It was not uncommon for early nineteenth century judges to dismiss suits on the ground that English precedent (which had been incorporated by statute into American law) would not allow the petitioners' result, while at the same time inventing new doctrines without consulting the legislature where (they claimed) precedent was silent.[138] In response to criticisms of the exercise of judicial power, defenders of the common law argued that reasoning from precedent was the best device for insuring that the law, if it were to change, would change in accordance to principle and logic.[139] Precedent was a constraint on discretion exercised either by legislative bodies or by judges.[140] For

[133] Cover, *Justice Accused*, n. 1 above, at pp. 135–43.

[134] Ibid., p. 27.

[135] Ibid., p. 137.

[136] See Cover, *Justice Accused*, n. 1 above, at pp. 140–1 and see Horwitz, *Transformation of American Law 1780–1860*, n. 95 above, at pp. 256–8 (on rise of codification movement and the emergence of the "formalist" picture of adjudication as will-less).

[137] The codifiers were certainly reacting to the legacy of the Marshall Court; as White noted, Marshall took it as a "political axiom" that constitutional law and common law were "inextricably linked" in early nineteenth-century jurisprudence: G. Edward White, *Marshall Court and Cultural Change*, n. 103 above, at p. 125.

[138] See Max Radin, *Handbook of Anglo-American Legal History* 217 (St. Paul, Minn.: West Publishing Co., 1936), p. 217 and James Willard Hurst, *The Growth of American Law: The Law Makers* (Boston: Little, Brown, 1950), p. 351.

[139] See Shannon Stimson, *The American Revolution in the Law*, n. 103 above, p. 144 ("Story's *Commentaries*—dedicated to John Marshall—are notable for their attempt to reproduce for American jurisprudence the English common lawyer's understanding of the detached and independent function of the Court . . . almost in passing, Story denies the right of juries to determine the law"). Stimson followed White in noting that although Story, Kent, and Shaw saw themselves as heirs to Marshall and Hamilton in their opposition to Jeffersonian codifiers, the techniques adopted by Story and Kent were more technical and formalistic than those of Shaw. See ibid., at p. 143 (citing G. Edward White, *The American Judicial Tradition* (New York: Oxford University Press, 1976), p. 43). Thus, Stimson concluded that compared to Story and Kent, Shaw's "adjudicatory techniques most closely followed those of Marshall": ibid.

[140] As Cover noted, "defenders of the common law faith" recognised that they had to explain why *judicial* changes in precedent were superior to *legislative* alterations of the same scope. Francis Lieber's reasoning was typical of the sort of argument used to rebut the call for codification: "In the

this reason Story rejected codification, preferring instead that "the common law should be left in its prospective operations . . . to meet the exigencies of society by the gradual application of its principles in the courts of justice to new cases".[141] Shaw, like Story, recognised that the *source* of common law was the precedents passed down from earlier courts, to which each judge was to apply principles of legal reasoning that were themselves learned through careful study of the common law. To put it in Hart's terminology, the rule of recognition for Shaw was not only the Constitution, but the rules of interpretation that allowed him to recognise common law precedents that were authorised under the Constitution.[142]

What distinguished Shaw, however, was his decision to return to *Somerset* and his insistence that its holding informed his reading of the Fugitive Slave Clause. It is difficult to see how Shaw could have been acting like a classical legal positivist in these instances, since it is difficult to see how *Somerset* could have been viewed as a command from the "sovereign" in America in 1836. The problem is not that, as a common law rule, *Somerset* had not been articulated in any piece of legislation. That would have left it in the same position as all of the common law, and the classical legal positivists had an awkward, but effective explanation as to how the large body of common law could be reconciled with the command thesis. Austin, for example, conceded that much of English law had never been "commanded" by Parliament but insisted that the *failure* of the Parliament to amend or reverse the decisions of the courts (which, even more so than in America, it had the power to do) was evidence of Parliament's tacit authorisation of the rules of law generated by the courts in their decisions.[143] The reason why it is difficult to reconcile *Somerset* with the command theory is that it had been in disuse for so long. It is very difficult to say that the American "sovereign" (the "People" named in the Constitution) had impliedly authorised *Somerset* when there was no evidence to that effect. While *Somerset* had not been repudiated by the Founders, it had not been applied with any rigour between 1787 and 1836. The same could be said for Shaw's interpretation of the

first place, precedent had a 'natural' power of its own. As a form of reasoned consistency, it has an attraction that is universal. Second, even the doctrine of precedent, which embraces change, requires a reason for departure from the old rule. And the necessity of giving a reason permits an analysis of the 'soundness' of prior judicial decisions. Thus, there is a basis for distinguishing or denying judicial decisions": Cover, *Justice Accused*, n. 1 above, at p. 143 (citing Francis Lieber, *Legal and Political Hermeneutics* (St. Louis: Filt, Thomas & Co. 1880), p. 185).

[141] Joseph Story, "Report on the Practicability And Expediency of Reducing to a Written and Systematic Code the Common Law of Massachusetts" in Jeremy Bentham, *Codification of the Common Law* (1882), p. 14.

[142] See Anthony J. Sebok, *Legal Positivism in American Jurisprudence* (New York: Cambridge University Press, 1998), pp. 306–7 citing Kent Greenawalt, "The Rule of Recognition and the Constitution", (1987) 85 *Michigan Law Review* 621 at 659 (Greenawalt's recognition that the rule of recognition in the USA could not be simply all or part of the Constitution but had to include "standards of interpretation").

[143] See John Austin, *The Province of Jurisprudence Determined* (H.L.A. Hart (ed.), 1954), pp. 30–2; H.L.A. Hart, *The Concept of Law*, n. 122 above, pp. 45–7; Joseph Raz, *Concept of a Legal System*, 2nd edn. (New York: Oxford University Press, 1980), p. 39.

Fugitive Slave Clause, which relied on his assumption that *Somerset* was law when the Clause was drafted, and that the Fugitive Slave Laws had to be read against a backdrop of *Somerset*. Again, Shaw was attributing a command to the sovereign for which there was simply no evidence of even a "tacit" command.

One of the benefits of abandoning the command theory and following Hart's social rule theory is that it allows the positivist to make sense of complex legal events like Shaw's adoption of *Somerset* and even his interpretation of the Fugitive Slave Clause in *Aves* and *Sims*. The key difference between the command theory and the social rule theory is the difference between a command and a practice. Practices are often more fluid and complex than commands, and can come into existence gradually.[144] Practices, no less than commands, are based in social sources, but those sources are not just the overt acts of specific legal officials, but the whole range of behaviours of the relevant legal officials. Since a social rule includes "a shared critical or reflective attitude towards the practice" in question, evidence of the existence of a social rule may include disagreement over the application of the rule, as long as the reasons for the disagreement reveal shared principles of interpretation.[145] Since the social rule theory does not view law as an idealised form of command, the need for the positivist to search for authorial intent in law is reduced. A legal rule is constituted by its meaning, not by what it was intended to do (although the rules of interpretation of a legal system may require that one take into account the intended purpose of a rule in the course of determining its meaning).[146] For all these reasons, Hart's social rule theory is far superior to the command theory in explaining how a judge approaches the tasks of applying abstract common law principles and constitutional principles.

From the perspective of Hart's social rule theory, Shaw's positivism becomes clear. While Shaw cared about the language of the Constitution, he cared because the rule of recognition for his legal system directed him to begin with the language. The rule of recognition, as Shaw understood it, also directed him to apply to the language complex principles of interpretation, which included, in part, the consideration of what the Constitution meant in light of its goals, structure, and the law that existed when it was written. Thus, Shaw felt compelled—by the rule of recognition—to inquire into the rules surrounding the conflict of laws and the status of *Somerset*. The fact that *Somerset* had not been adopted and seemed to have been disfavoured since the founding was not, in itself, a sufficient reason to reject it; as Hart wrote, "any honest description of the use of precedent in English law must allow" for certain "contrasting facts", such as the fact that although certain rules can be "extracted" from past decisions, those rules may be isolated or even reversed depending on further review of the very precedents from which the rule was drawn.[147] Shaw was obliged to

[144] See Hart, *Concept of Law*, n. 122 above, at p. 262.
[145] Ibid., pp. 265–6.
[146] Ibid., pp. 252–3.
[147] Ibid., pp. 134–5 and 268 (on the identification of "latent legal principles" in common law).

take into account *all* of the sources of law relating to the problem of slaves in transit: recent Massachusetts decisions, other state decisions, English decisions including *Somerset*, and the principles of interpretation that were accepted by the legal officials in Massachusetts, including any principles that might have related to the "weight" of competing legal principles.[148] Just as Justice Cardozo argued that a principle, heretofore unarticulated or recognised in New York law, outweighed competing sources of decision (such as certain methods of statutory interpretation) in *Riggs v Palmer*, so Shaw determined that the principle instantiated by *Somerset* outweighed competing social sources of decision (such as related precedent, policy concerns, and statutes from neighboring states).[149] Thus, Shaw's decision to adopt *Somerset*, which, like his private decisions, might have struck Nelson as an act of sheer instrumentalism, was consistent with a form of positivism that adopted a broad understanding of the rule of recognition for his legal system. Similarly, his interpretation of the Fugitive Slave Laws, which struck Cover as a formalistic application of precedent without regard to principle, was also consistent with a form of positivism that measured precedent, the language of the Fugitive Slave Clause, and the common law against the best rational construction that could be imposed upon the Constitution as a whole.

VIII. THE BURDEN OF POSITIVISM

In conclusion, let me stress that the point of my argument is not to prove that Shaw's decision in either *Aves* or *Sims* was correct. I do believe that it is unlikely that Shaw could have faithfully decided for liberty in both cases. I think that Shaw was right to conclude in 1836 that if he adopted *Somerset* on behalf of Massachusetts, he was committed to interpreting the Fugitive Slave Clause as a grant of exclusive federal power. Seven years ago I presented a "Dworkinian" argument for the view that the Fugitive Slave Clause was incompatible with the Fugitive Slave Act of 1850.[150] The only problem with that argument, which may still be valid, is that it assumed that the Constitution guaranteed comity in matters concerning slavery—hence, it implicitly assumed that *Somerset* was inapplicable in antebellum America.[151] The argument I am making here is based on the rejection of that assumption. It takes as its reference point the idea that

[148] There is no reason why a principle cannot comprise the content of a social rule: "arguments from . . . non-conclusive principles are an important feature of adjudication and legal reasoning, and . . . should be marked by an appropriate terminology" by the positivist. Hart, *Concept of Law*, n. 122 above, at 263. Hart admitted that he was not as clear as he should have been when he first wrote *The Concept of Law* on the place of principles in his social rule theory: "I certainly did not intend my use of the word 'rule' to claim that legal systems comprise only 'all-or-nothing' or near-conclusive rules": ibid.

[149] See Hart, ibid., p. 262 (Dworkin was wrong to conclude that Cardozo's views on *Riggs v Palmer*, 115 N.Y. 506 (1889), could not be explained by Hart's legal positivism).

[150] See Anthony J. Sebok, "Judging the Fugitive Slave Acts", (1991) 100 *Yale Law Journal* 1835.

[151] Ibid., at 1848.

Somerset was correctly made part of American law by Shaw. Consequently, one goal of this essay is to characterise the legal reasoning that led Shaw to adopt *Somerset*. As I have argued above, Shaw's reasoning reflected a sophisticated form of legal positivism. Furthermore, another goal of this essay is to show that *Aves* and *Sims* were part of a single coherent vision of the relationship between slavery and the law. That is, that the reasoning that led Shaw to free Med in *Aves* was deeply connected to the reasoning that led him to send Sims back to bondage. Thus, *Sims*, no less than *Aves*, was rooted in a sophisticated form of positivism.

It is not my purpose in writing this essay to choose between the Dworkinian model I developed seven years ago and the Hartian model I develop in this article. My purpose is to defend Shaw against his critics. Shaw's legal reasoning was neither formalistic nor instrumental: it was based on a careful attempt to make sense of a complex body of law in which the law's clear commitment to liberty for slaves in transit was the best evidence that the Constitution gave the power to regulate the capture of fugitive slaves to the federal government. It is possible that Shaw did not have to embrace *Somerset* as fully as he did. But given his decision to decide *Aves* by incorporating *Somerset*, it is difficult to see how he could have decided *Sims* without following his interpretation of *Somerset* to its logical conclusion—the Fugitive Slave Clause. If, as I have argued, the decision to adopt *Somerset* was neither formalist nor instrumentalist, it is not clear why the decision to carry forward that adoption became formalist or instrumentalist simply because the outcome was unjust. Once we understand why Shaw approached the Fugitive Slave Clause the way he did, then we must admit that his final conclusion in *Sims*, although questionable, was based on a form of legal reasoning that was sophisticated in its method and broad in its scope. Thus, an important lesson of this study of Shaw's interpretation of antebellum slave laws is that although positivism (in the hands of a good judge) can improve the quality of adjudication in a given legal system, it can never insure against an unjust result.[152]

[152] See Sebok, *Legal Positivism in American Jurisprudence*, n. 142 above, at pp. 7–17 (arguing that Sophocles's Creon—who was a positivist and a good judge—correctly applied bad laws and therefore produced tragic results).

7

The Rule of Law and Judicial Review: Reflections on the Israeli Constitutional Revolution

ALON HAREL*

I. INTRODUCTION

The rule of law is a theoretical concept discussed extensively by philosophers and legal theorists. At the same time, the rule of law is a legal concept used by lawyers and judges. The meaning of this concept should be informed by both moral and legal theorists who investigate the abstract moral and political aspirations embodied in the concept of the rule of law and by legal practitioners who transform these aspirations into legal reality.

But as many theorists discussing the rule of law have realised, the rule of law is not simply a set of mechanical rules to be followed. It is a set of loose, vague and indeterminate principles which require interpretation in light of the values which the rule of law is designed to realise. Fuller has imaginatively articulated this point by stating that the rule of law "is condemned to remain largely a morality of aspiration and not of duty".[1]

Fuller's observation suggests that the implementation of the principles of the rule of law depends upon the ways in which it is viewed and interpreted by courts. This essay is devoted to exploring the ways in which Israeli courts understand and interpret the concept of the rule of law. More particularly, it explores whether the principles of the rule of law—principles whose meaning is not self-evident—can be implemented faithfully by courts which face an institutional crisis. Its primary conclusion is that the Israeli Supreme Court errs in its claim that judicial review is necessarily conducive to the protection of the rule of law. It also shows that the concept of the rule of law as understood by the Israeli Supreme Court serves the Court's institutional purposes including the purposes

* I wish to thank Brian Bix, Celia Fassberg, Ruth Gavison, Doron Kalir, Michael Mandel, Andrei Marmor, Mike Otsuka and Ed Rock for their important comments on earlier drafts. I am also grateful to Danny Priel who provided excellent research assistance.

[1] See Lon L. Fuller, *The Morality of Law* (New Haven: Yale University Press, revised edn., 1969), p. 43.

of expanding the Court's powers, legitimating its controversial role in Israeli constitutional law and establishing the Court's privileged position in Israeli constitutional discourse.

The investigation conducted in this article does not exhaust the multiplicity of meanings attributed to the term 'the rule of law' in Israeli jurisprudence. Instead, I chose to investigate the ways in which the Court uses the ideal of the rule of law to justify and protect its constitutional powers, in particular, the power of judicial review. The justification for limiting the discussion in this way is grounded in the special constitutional importance of the debate over judicial review in Israel—a debate which currently dominates Israeli constitutional discourse.

The essay is divided into two main parts. Section II explores the recent changes in Israeli constitutional law, in particular, the rise in the powers of the Court and the controversies concerning these changes. It also explores the ways in which the Court justifies its newly acquired powers, in particular, its claim that its constitutional powers are conducive to the protection of the rule of law. Section III examines critically the Court's use of the ideal of the rule of law to justify judicial review. More specifically, it exposes the potential destructive impact of the Court's understanding of the concept of the rule of law on Israeli constitutional discourse.

II. THE RULE OF LAW IN ISRAEL

1. Israeli courts: the new institutional realities

Israeli courts have become in recent years both very powerful and activist courts.[2] There are two primary phenomena which illustrate this development: changes in judicial reasoning and rhetoric, and changes in constitutional legal doctrine.

The reasoning of the courts changed from a formalistic one to reasoning which relies more explicitly on moral and political values.[3] Abstract values such as equality, freedom and justice are used more often by the courts while doctrinal "black letter" type of arguments have become less prevalent. The extensive

[2] The terms "activist" and "powerful" are different, but in the Israeli context the activism of the Court and its power are interrelated. The power of the Court in Israel is based on its own understanding and interpretation of Israeli constitutional law. Activist decisions by the Court expanded its powers to interfere in the decisions of executive agencies and of the legislature. The greater powers acquired by the Court as a result of this interpretation, in turn, facilitate greater activism on its part.

This essay focuses on constitutional law. However, the activism of the Court is reflected also in its decisions in the area of administrative and private law. For instance, courts have recently become more inclined to interfere in contracts in order to bring about just outcomes.

[3] Menachem Mautner, *Decline of Formalism and the Rise of Values in Israeli Law* (Tel Aviv: Ma'agalay Da'at, 1993).

use of abstract moral and political values in legal reasoning is accompanied with greater inclination on the part of judges to interfere in the executive's decisions.[4]

But the changes are not merely changes in the forms of legal reasoning and rhetoric. Recent changes in constitutional law have also increased the powers of the Court. The Israeli Supreme Court narrowed the scope of doctrines that limited the rights of petitioners to challenge the executive's decisions.[5]

The most important change explaining the rise in the powers of the Court is the enactment of two basic laws in 1992: Basic Law: Human Dignity and Liberty and Basic Law: Freedom of Occupation.[6] The two basic laws are in the view of

[4] It is not claimed here that the changes in reasoning *explain* the greater activism on the part of the Court. It is more accurate to say that the changes in the inclinations of the Court explain the changes in the reasoning and that the changes in the reasoning reinforce, in turn, the activist inclinations of the Court.

[5] The two primary examples are the changes in the doctrine of standing and the political question doctrine. See *Ressler v Minister of Defense* 42(2) PD 441 (1988). For an English translation of Ressler, see (1996) 10 *Selected Judgments of the Supreme Court of Israel 1988–1993* 1.

[6] Let me provide some background concerning these laws. Basic Law: Human Dignity and Liberty asserts in section 1 that "Fundamental human rights in Israel are founded upon recognition of the value of the human being, the sanctity of human life, and the principle that all persons are free". Section 2 prohibits "violation of the life, body or dignity of any person as such". In addition to the abstract rights enumerated in this section, the Basic Law protects more specific rights such as the right to property (section 3), the right to exit the country (section 6(a)), the right of citizens to enter the country (section 6(b)), the right of privacy (section 7(a)) and the right to personal liberty (section 5). The Basic Law however qualifies the protection of rights in section 8 as follows: "There shall be no violation of rights under this Basic Law except by a law befitting the values of the State of Israel, enacted for a proper purpose, and to an extent no greater than is required". This qualification is often compared with section 1 of the Canadian Charter which subjects the protection of rights to "such reasonable limits prescribed by law as can be demonstrably justified in a free and democratic society".

The interpretation of Basic Law: Human Dignity and Liberty generated numerous disputes, in particular, disputes concerning the scope of the rights protected in section 2. The Court has a very expansive interpretation of the term "dignity"—an interpretation which is disputed by some legal scholars. See n. 10 below.

Basic Law: Freedom of Occupation protects in section 3 the right of any Israeli national or resident to "engage in any occupation, profession or trade". The history of this Basic Law can illustrate the vulnerability and fragility of Israeli basic laws. Basic Law: Freedom of Occupation was first enacted in 1992. However, shortly after its enactment in 1994, Basic Law: Freedom of Occupation was repealed and reenacted with some important modifications. This peculiar event was a byproduct of a political crisis generated by the decision of the Court under which the prohibition on the importation of non-kosher meat violates the right protected in section 3 of Basic Law: Freedom of Occupation. In order to overcome the decision of the Court, the new version of Basic Law: Freedom of Occupation included in section 8 a provision similar to section 33 of the Canadian Charter under which the Parliament can override the restriction by enacting a law "passed by a majority of the members of the Knesset, which expressly states that it shall be of effect, notwithstanding the provisions of the basic law". Then the Parliament passed a law which explicitly prohibits the importation of non-kosher meat satisfying the conditions set up in section 8. A petition challenging the constitutionality of this new law was rejected by the Court and the stability of the coalition was saved.

For a thorough survey of the history of these constitutional changes, see Ruth Gavison, "The Constitutional Revolution: A Reality or Self Fulfilling Prophecy" (1997) 28 *Mishpatim* 21 at 95–110. For surveys in English, see Daphne Barak-Erez, "From an Unwritten to a Written Constitution: The Israeli Challenge in American Perspective" (1995) 26 *Columbia Human Rights Rev*. 309; Stephen Goldstein, "Protection of Human Rights By Judges: The Israeli Experience" (1994) 38 *Saint Louis University L. Rev*. 605; David Kretzmer, "The New Basic Laws on Human Rights: A Mini-Revolution in Israeli Constitutional Law?" (1992) 26 *Isr. L. Rev*. 238.

For an English version of Basic Law: Human Dignity and Liberty, see <http://www.israel.org/

many the first serious attempt by the legislature to establish a bill of rights by legislating a series of "basic laws" that would enjoy priority over ordinary statutes legislated by the Parliament.

Some jurists, including the Israeli Chief Justice Aharon Barak, described these basic laws as "a constitutional revolution". Under this view, the enactment of these statutes is a critical step in a process by which Israel abandons its British constitutional conception of unlimited sovereignty of Parliament and moves into an American-style system under which a written constitution limits the powers of the Parliament.[7]

The view that the basic laws change the Israeli constitutional structure was adopted by the Court in one of the most important cases in Israeli constitutional history. In *Bank Hamizrachi v Migdal*, the Court discussed at great length the scope of its own powers and declared unambiguously that it has the power to invalidate statutes which conflict with these basic laws.[8] This 368-page decision can only be compared in its importance to the case of *Marbury v Madison* in the American system.

While the Court in this case failed to use the power it declared (at great length) it had, it did not hesitate to use its newly acquired constitutional powers in a later case.[9] Its willingness to use this power motivates lawyers to petition the Supreme Court on constitutional grounds and influences the legislature in its deliberations concerning new legislation. The constitutional powers of the Court are therefore an important factor in Israeli legal and political life.

gov/laws/dignity.html>. For an English version of Basic Law: Freedom of Occupation, see <http://www.israel.org/gov/laws/freeoccu.html>.

[7] Claude Klein was the first to announce the existence of a constitutional revolution after the enactment of the two basic laws in his article: "The Quiet Constitutional Revolution", *Ma-ariv*, 27 March 1992. The name Klein chose is borrowed from the quiet revolution which took place in Quebec in the 1960s.

Chief Justice Aharon Barak adopted the term. See Aharon Barak, "The Constitutional Revolution: Protected Human Rights" (1992) 1 *Mishpat Umimshal: Law and Government in Israel* 9. In the third volume of his monumental treatise on interpretation, Barak makes the following controversial statement: "Israel is a constitutional democracy . . . The constitution—in the image of the protected basic laws is the ultimate legal norm". See Aharon Barak, *Interpretation in Law* (Jerusalem, Nevo Press, 1994), vol. III, pp. 37–8. Barak restated this view in *Bank Hamizrachi v Migdal* 49(4) PD 221, 352. For references describing these debates, see n. 10 below.

[8] *Bank Hamizrachi v Migdal* 49 (4) PD 221 (1995)

[9] See *Organization of Investment Consultants v The Minister of Finance* 97(3) Tak Elion 721 (1997). Michael Mandel described the case as a case of "monumental triviality": see Michael Mandel, "The Legalization of Politics in Israel" (unpublished manuscript). A group of stockbrokers complained that the transitional provisions of the new law to regulate the stock market exempted only those with seven years' experience from having to take competency tests. The requirement to take competency tests, it was claimed, violated the rights of stockbrokers to freedom of occupation. The Court balanced carefully the rights of the stockbrokers against the competing societal considerations and declared with great passion that seven years is too much! The Knesset saved the fundamental rights of stockbrokers by reducing the period to five years! This case illustrates the complexity of the political considerations facing the Court in using the basic laws to invalidate legislation. On the one hand, using the basic laws to strike down trivial provisions of this type may raise the concern that the rhetoric of rights is devalued. On the other hand, for political reasons, the Court cannot use the basic laws to strike down some of the more blatant violations of human rights since such a decision may lead to a direct confrontation between the Court and the executive or legislature.

Israel is hardly the only country in which such a transition has taken place. Another primary example is Canada which adopted in 1982 a Charter which provides constitutional protection of various rights. Israel differs however from Canada in that the Israeli "constitutional revolution" is primarily guided by the courts, in particular, the Israeli Supreme Court. Although certain powers of judicial review were given to the judges in the two basic laws, the scope of these powers is controversial and the courts have a key position in interpreting their powers under the basic laws.

The constitutional revolution is bitterly contested in Israel.[10] Its opponents can be classified into two types. The first type of opponents are conservative or traditional opponents. Their opposition is founded on the belief that Israeli courts are fundamentally liberal and Western and consequently expanding their powers threatens the Jewish foundations of the State.[11] Their resentment can be better understood if one remembers that the rise in the political power of the courts coincides with the recent rise in the political power of religious and Sephardi parties. A cynic may infer that the courts are used by the elite in its struggle against the recent rise of new conservative political powers.[12] Ironically, however, the constitutional revolution has generated a second type

[10] The controversies range over many different issues. One issue is identifying the rights protected by the term "dignity" in Basic Law: Human Dignity and Liberty. Some believe that the term dignity should be interpreted broadly to include the right to free speech, the right to equality and numerous other rights which are not specified in the basic laws. The rights to freedom of speech and equality are already protected in Israel by the courts but interpreting the term "dignity" in the broad way suggested by the judges would grant these rights constitutional protection. See Hillel Sommer, "The Non-Enumerated Rights: On the Scope of the Constitutional Revolution" (1997) 28 *Mishpatim* 341. In addition, there are controversies as to the powers of the courts to invalidate statutes and the powers of the legislature to override basic laws. See Yoav Dotan, "Constitution to Israel?—The Constitutional Dialog After 'The Constitutional Revolution'" (1997) 28 *Mishpatim* 149.

The broad interpretation of the powers of the Court was criticised by the opponents of the constitutional revolution including former Chief Justice Moshe Landau. See Moshe Landau, "The Supreme Court as Constitution Maker for Israel" (1996) 3 *Mishpat Umimshal: Law and Government in Israel* 697; Moshe Landau, "Reflections on the Constitutional Revolution" (1996) 26 *Mishpatim* 419 . See also Ruth Gavison, n. 6 above, pp. 27–33.

The central role of the courts in this process explains the rhetoric used by the opponents of the "constitutional revolution". These opponents do not criticise a revolution which they concede took place, but rather dispute the very reality of a "constitutional revolution". More specifically, they argue that the Court declared a revolution which has never taken place. See Gavison, ibid., at 129–32.

[11] Most forcefully this view was expressed by Arye Deri, an orthodox member of the Parliament who says: "These three proposed innocent laws which . . . my humble mind cannot find any faults in them . . . But who interprets these laws . . .? The Supreme Court interprets them and in its contemporary composition and its liberal interpretation . . . I cannot as an ultra-orthodox Jew . . . rely on the interpretation of the Supreme Court . . . Even if the Ten Commandments were brought as a proposed Basic Law I would oppose them. . . I do not know what you conspire with the Supreme Court judges to do": cited in Dotan, n. 10 above, at 196 n. 88. For a description of the conservative opposition to the Court, see Ronen Shamir, "The Politics of Reasonableness: Reasonableness and Judicial Power at Israel's Supreme Court" (1995) 5 *Theory and Criticism* 7.

[12] See Mandel, n. 9 above.

of counter-revolutionaries: the liberal counter-revolutionaries.[13] The liberal counter-revolutionaries oppose the constitutional revolution for various reasons. Some express majoritarian concerns and argue that it is improper that the Court declares a constitutional revolution whereas such a change was not envisaged by the legislature and is not supported by the public.[14] Others believe that a written constitution may lead to greater caution on the part of the judges and therefore may eventually be detrimental to the protection of human rights.[15] This conviction is supported by the widespread belief that the violation of human rights in Israel is not the byproduct of weakness in legal doctrine. In fact, Israeli judges who wanted to protect certain rights such as the right to free speech were very innovative in creating powerful interpretative tools to protect the rights they favoured. The Court's failure to protect some other fundamental rights cannot therefore be attributed to the poverty of the legal doctrine, but to the Court's unwillingness to protect these rights.[16] Most importantly, however, liberal counter-revolutionaries are convinced that a judicial constitutional revolution exposes the courts to political pressures. Courts which have to defend their very constitutional powers to review statutes are less able vigorously to defend human rights. Ironically, under this view, it is the Court's declaration of de jure constitutional power which weakens its de facto power and consequently undermines its ability and commitment to the protection of human rights.[17]

[13] Some opponents of the constitutional revolution are hard to classify as either liberal or conservative. Thus, for instance the position taken by former Chief Justice Moshe Landau is based on his understanding of the judicial role and does not stem from any liberal or conservative ideology. See Landau, n. 10 above.

There is a difference in the way judges react to the conservative and the liberal opposition. Judges are often more concerned to address the liberal opposition because the discourse of liberal opponents of the courts is often phrased in legal language and is made by senior members of the Israeli legal community. The conservative opposition is conducted in different language and is addressed to different communities.

[14] See Gavison, n. 6 above at 70, 126. Gavison emphasises the need for a public acceptance of the constitution. See Gavison, ibid. at 70–2. She also shows that there was no intention on the part of the legislature to adopt a constitution, that the judicial creation of a constitution is elitist and that the absence of clear shared views concerning the content of the constitution is divisive. Gavison, ibid. at 115–22. This position is also held by Justice Cheshin who opposes the label "constitutional revolution" to describe the two basic laws. See *Bank Hamizrachi*, n. 8 above, at 522–5. Mandel provides a Marxist perspective under which the Israeli Court represents secular established factions in the Israeli society in their political struggle against traditionalist factions. The Court's interference is therefore anti-democratic in his view because it aims at protecting established elites against the emergence of new popular political powers. See Mandel, n. 9 above.

[15] See Leon Sheleff, *The Rule of Law and the Nature of Politics* (Papyrus: Tel Aviv University, 1996), p. 63.

[16] See Mandel, n. 9 above. It is worthwhile to note in this context that even in the USA there is less reliance on the Constitution in protecting human rights and greater reliance on statutes. See Henry Paul Monaghan, "The Age of Statutory Fundamental Rights" (1993) XIV *The Toqueville Review* 139. Monaghan believes that the USA has entered the "Age of Statutes" and that the age of the Constitution is over: see ibid. at 145.

[17] This claim is based on a speculation but some of the recent decisions of the Supreme Court suggest that this speculation is not unfounded. The Supreme Court has been effective in some areas such as protection of speech and assembly, women's equality and gay rights; but it miserably failed in the

Thus, many Israeli liberals and human rights activists found themselves in a peculiar position. While advocating a more activist role for the Court in defending human rights, in particular, on issues such as administrative arrests, house demolition, practices of torture used by the Israeli security services etc., they also argue against the de jure expansion of judicial powers, i.e. against the constitutional revolution.

2. The rule of law and judicial review

A complete investigation of the concept of the rule of law as understood by judges requires not merely an understanding of the judicial abstract declarations concerning this concept, but an examination of the contexts and issues in which judges use the term 'the rule of law' as well as the contexts in which they refrain from using it.[18]

The purposes of this essay however are more limited. My aim is to investigate the claim that judicial review is conducive to the rule of law. In numerous decisions, the rule of law is perceived by the Court to require a certain division of labour between the three branches of government. In some of these decisions, the rule of law limits the extent to which the legislature can delegate its power to various agencies.[19] In other decisions, the Court uses the term to justify a narrow construction of the discretionary powers of the executive agency.[20] But the most important use of the term the rule of law in this context is to justify the constitutional powers of the Court, i.e., to justify the institution of judicial review.

most important and crucial areas, namely the protection of Palestinian rights. See Andrei Marmor, "Judicial Review in Israel" (1996) 4 *Mishpat Umimshal* 133 at 137–9. It is naturally hard to know what the precise reasons for this failure are, but there is some basis for the claim that the Court is under severe political pressures since the constitutional revolution and consequently is limited in the degree to which it can protect human rights.

This observation does not entail that in the absence of a constitutional revolution, the Court would have protected vigorously rights of Palestinians. The record of the Court in protecting Palestinian rights had been poor even before the basic laws were enacted. See, e.g. *Association of Civil Rights v The Minister of Defense* 47(1) PD 268 (1992). In this case, the Court upheld the decision to expel 400 Palestinians without due process. For a powerful critique, see Eyal Benvenisti, "Judicial Review of Deportation Orders" (1993) 1 *Mishpat Umimshal: Law and Government in Israel* 441. However, it is claimed that the constitutional revolution made it even more difficult for the Court to protect these rights.

[18] It is easier to analyse the concept of the rule of law on the basis of the cases in which judges use the term. However, it is equally important to explore in addition contexts in which judges refrain from using the term. The petitions concerning torture are perhaps a prime example. The torture conducted by the security services in Israel is not authorised by the law and could therefore be naturally described as violation of the rule of law. But the Court refrained from using this terminology in this context.

[19] See *Shansi v The Supervisor of Commerce in Diamonds*, 97 (3) Tak-Elion 476, 488.

[20] See, e.g., *Schnitzer v The Chief Military Censor* 42 (4) PD 617, 634–7. An English translation of the case is also available. See (1995) 9 *Selected Supreme Court Judgements* (1977–1990) 77, 99–103.

The argument that judicial review is necessary for protecting the rule of law is not new to Israeli constitutional lawyers. It was first made by Professor Barak shortly before his appointment to the Supreme Court in an article published in 1977. In a somewhat revised form, it appears in Barak's most important judicial decisions.

The argument appears in two different versions. Under the earlier version, judicial review is necessary for the protection of the rule of law because:

> "The principle of the rule of law denotes a normative principle. In order to be governed by it, one needs to endorse it as a legal norm. But it is not sufficient that it be entrenched in a statute or bylaws. It has to be entrenched as a supreme legal norm such that the legislature cannot infringe it . . . This can be done by using a written constitution in light of which the legality of the decisions of the legislature, the executive and the judiciary can be examined . . . For the sake of promoting [the rule of law] one needs to establish a system of judicial review".[21]

This argument supports judicial review on the grounds that the principles of the rule of law need to be honoured not only by the executive, but also by the legislature. The legislature can violate the rule of law by enacting statutes which are vague, retroactive, contradictory, or violate in other ways the principles of the rule of law. Judicial review is required in order to protect against infringements of the rule of law by the legislature.

This argument is, however, limited in scope since it is only the principles of the rule of law themselves which, under this view, should benefit from constitutional protection. The rule of law does not require therefore the constitutional protection of human rights, or other important values.[22]

Once Israel adopted its basic laws in 1992, Barak developed a new argument premised on his belief that the basic laws impose limits on the powers of the legislature. Chief Justice Barak believes that:

> "When in a given legal system, there exists a constitution, the 'rule of law' requires maintaining the authority of the Constitution. Indeed, the Parliament by using its constituent authority provided the State basic laws. These are supreme norms in the normative hierarchy. In order to fulfill its will, it is necessary to invalidate statutes which conflict with basic laws . . . Indeed in declaring the invalidity of a statute which conflicts with the basic law, the Court applies the basic law. The legitimacy of the Constitution and the basic law legitimize judicial review of statutes . . . It follows that judicial review is the spirit of the Constitution. Without judicial review the Constitution has no life. Constitutional supremacy needs to be accompanied by judicial review".[23]

[21] See Aharon Barak, "The Rule of Law" in Shimon Shetreet (ed.), *Collection of the Lectures Delivered at the Seminar for Judges 1976: Recent Developments in Israeli Law and Legislation* (Jerusalem: The Faculty of Law, 1977), p. 24.

[22] This limit is not as significant as it may seem. Barak has a very expansive conception of the rule of law and, therefore, under his view, many basic rights constitute parts of the rule of law. See Barak, ibid. I am grateful to Danny Priel for raising this point.

[23] *Bank Hamizrachi*, n. 8 above, at 420.

Under this argument, when a legal system constrains the legislature, it needs to establish an institution which has the power to declare statutes invalid. Constraints on the legislature are meaningless unless they are accompanied by institutions which have the power to invalidate statutes which are incompatible with these laws. If courts do not have such powers, the legislature can violate constitutional provisions and consequently can act in a way which is detrimental to the rule of law.

The second argument is in my view a much more radical one in that it facilitates broader scope to judicial review. The first argument provides at most a justification for protecting the principles of the rule of law. The second argument provides a justification for protecting a full fledged bill of rights since in Justice Barak's view, Basic Law: Human Dignity and Liberty is a mini constitution and it provides constitutional protection to a broad set of human rights.

III. JUDICIAL REVIEW AND THE RULE OF LAW: A CRITIQUE

This section is divided into two sub-sections. The first sub-section illustrates that the Court's claim that judicial review is conducive to the rule of law is unfounded. More particularly, it shows that although judicial review promotes the rule of law in one respect, it may be detrimental to the rule of law in another respect. The second sub-section draws attention to the fact that justifying judicial review on the grounds that it is conducive to the rule of law may stifle political deliberation over some of the most important constitutional questions facing Israeli society.

1. Is judicial review conducive to the rule of law?

Justice Barak's argument favouring judicial review is based on two premises. First, he believes that the rule of law in its most basic and narrow meaning requires that individuals as well as institutions obey the law and be ruled by it. Second, he argues that the basic laws in Israel impose constraints on the powers of the Parliament. From these premises it follows that the rule of law will be compromised if the legislature does not honour the limits on its own powers imposed by the basic laws. Hence, Chief Justice Barak concludes, judicial review is necessary in order to protect the rule of law.[24]

[24] Naturally from the fact that judicial review reduces the risk of violation of the law by the legislature, it does not follow that Israel should have a system of judicial review. Arguably granting a talented and reliable constitutional scholar the power to invalidate statutes of the Parliament may also reduce the risks of enacting invalid laws and yet nobody would be inclined to adopt such a system. The rest of this section will however avoid this challenge and examine merely whether judicial review in fact reduces, as Justice Barak presupposes, the risks of institutional decisions which violate the law.

I have no dispute with any of the premises of Justice Barak. The rule of law requires that the legislature conforms with the constitutional constraints imposed upon it. The basic laws in Israel impose such constraints on the powers of the Parliament. But these premises, it will be argued, do not entail the conclusion favoured by Chief Justice Barak.

The rule of law is compromised whenever the legislature uses powers it does not have, i.e. enacts statutes which conflict with the basic laws. Judicial review is conducive to the rule of law in that it provides an opportunity to override the legislature's erroneous decisions. But this does not entail that judicial review is conducive to the rule of law *all things considered*. In order to evaluate this stronger hypothesis, one needs to evaluate whether judicial review may, in addition, be detrimental to the rule of law.

The opponents of judicial review would argue that judicial review creates the danger that the Court unjustifiably invalidate statutes which do not conflict with the basic laws. This possible danger could naturally be phrased as risking or endangering the rule of law. Thus, while it is true that judicial review is conducive to the rule of law because it facilitates the invalidation of statutes which conflict with the basic laws, judicial review may also be detrimental to the rule of law when the Court invalidates statutes which do not conflict with the basic laws.[25]

Can Chief Justice Barak salvage the validity of his argument? The validity of this argument could be salvaged by adding an additional premise (the greater risk premise), namely the premise that the first type of risk—the enactment of an unconstitutional law is a more serious risk to the rule of law than the latter type of risk, namely that the Court erroneously invalidates statutes which do not conflict with the basic laws.

Justice Barak can provide three justifications for the greater risk premise. Under the first justification, the Court is indeed a more reliable interpreter of the Constitution because of its special institutional position. The Court, under this

[25] This is not a new argument. Thomas Jefferson—the most well know opponent of judicial review—made this point in 1819. In his letter to Judge Spencer Roane, Thomas Jefferson criticised judicial review on the grounds that: "The constitution, on this hypothesis, is a mere thing of wax in the hands of the judiciary, which they may twist and shape in any form they please . . . My construction of the constitution is very different from that . . . It is that each department is truly independent of the others, and has an equal right to decide for itself what is the meaning of the constitution in the cases submitted to its action; and especially, where it is to act ultimately and without appeal": see letter to Judge Spencer Roane, 6 September 1819. Reprinted in Saul K. Padover, *Thomas Jefferson and the Foundations of American Freedom* (Princeton: New Jersey, 1965), pp. 162–3. I am grateful to Eyal Benvenisti for drawing my attention to this letter.

In Israel this argument was raised by Marmor, n. 17 above, at 152–9. Marmor points out the middle class values shared by judges—middle class values which do not necessarily reflect heterogeneity of the Israeli society; the fact that decisions concerning constitutional rights are not merely legal questions in a technical sense but moral issues, and lastly the fact that there is no consensus concerning the scope of constitutional rights. Consequently Marmor believes that the Supreme Court has no clear advantage over political bodies in making decisions concerning the constitutionality of statutes.

argument, is a more objective and reliable institution than politicians and consequently the risks of an unjustified enactment and enforcement of rules which conflict with the basic laws are greater than the risks of an unjustifiable act on the part of the judiciary.[26]

This claim, however, must be based on an evaluation of the specific societal background. One can imagine a society where the political discourse requires politicians to be faithful to constitutional principles because political discourse is conducted in constitutional language. In these societies, the legislature may remain loyal to constitutional principles without judicial review because such a loyalty is dictated by the legislature's own interests. Thus, Chief Justice Barak's belief that courts are more reliable interpreters of the basic laws is not a universal argument; instead it depends on the particularities of the Israeli society and its institutional dynamics. Barak, however, fails to provide reasons based on the particularities of the Israeli society which support his conclusion.

I do not claim however that judicially unenforceable rights can effectively be protected in the Israeli context. Israel is not a society in which the protection of rights constitutes an important part of the political discourse. This point illustrates, however, that there is nothing incoherent, or unintelligible in a system restricting the powers of the legislature without establishing an institution of judicial review. The desirability of judicial review depends largely on the specific cultural and political context. As I shall show later, the relevant circumstances should include the past performance of the Court itself and without evaluating its relative success, no firm conclusions as to the desirability of judicial review can be made.

The second justification for the greater risk premise is based on the belief that errors on the part of the legislature, namely errors which result in the enactment and enforcement of invalid laws are more costly than errors on the part of the judges, namely errors which result in the invalidation of valid laws. A possible justification is grounded in the claim that the basic laws in Israel protect human rights and, arguably, it is always better to make an error which protects unjustifiably a claim which should not be protected than to err by violating unjustifiably a legal right which should have been honoured. Judicial review creates therefore a risk which is worth taking given the alternative feasible institutional mechanisms.

Such an argument can be criticised on several grounds. First, the basic laws do not protect every important human right. Hence, a conflict could arise between a right which is protected by the basic laws and a right which is not protected by them. A wrong decision by the Court to invalidate a statute could therefore violate a right. Secondly, it is false to presuppose that it is always better to expand unjustifiably the scope of a constitutional right than to fail to protect a right which should be protected. The risks of an error cannot be determined in abstract. They should be determined in light of the importance of the conflicting rights and interests involved.[27]

[26] See Aharon Barak, *Interpretation in Law*, n. 7 above, at p. 109.

But even if the view that the costs of a wrong decision made by the Court are lower than the costs of a wrong decision made by the legislature is correct, this view fails to prove the greater risk premise. Under the greater risk premise, the enactment of an unconstitutional law is a more serious risk *to the rule of law* than the risk that the Court erroneously invalidates statutes which do not conflict with the basic laws. Judicial review may be desirable on the grounds that it promotes the protection of human rights and yet not be conducive to the promotion of the rule of law.[28]

The protection of the rule of law is often regarded as conducive to the protection of human rights. Thus, for instance, Raz believes that the protection of the rule of law is necessary to the protection of values which are often central also to the justification of rights such as human dignity,[29] or freedom.[30] But while protecting the rule of law is often conducive to the protection of human rights, the two concepts are not identical; moreover one can easily conceive of measures which would promote the one at the expense of the other. Thus, an executive body committed to the protection of human rights which would consistently frustrate the enforcement of laws which violate human rights may promote the protection of human rights at the expense of the principles of the rule of law. Hence, illustrating that judicial review is conducive to the protection of human rights may be a powerful argument favouring judicial review but it is not sufficient to show that it is conducive to the protection of the rule of law.

Sometimes Justice Barak seems to support the greater risks premise on the grounds that the Court has the authority to interpret the basic laws. For instance, Justice Barak states that: "In a democratic regime, based on separation of powers, the authority to interpret laws including basic laws . . . is in the hands of the Court".[31]

This argument is based on two premises. First, it is based on the conviction that the Court has the power to interpret statutes. Secondly, it presupposes that judicial review is primarily a matter of interpretation of the basic laws. Consequently, judicial review is simply a natural extension of the judicial role, in particular, its role in interpreting statutes.

This argument fails because it presupposes what needs to be proven. Naturally courts are often engaged in interpretation of the law. But the authority to interpret the law is not a premise which justifies the power of the Court. Instead the authority to interpret the basic laws is a by product of an indepen-

[27] The case of L*ochner* in which the United States Supreme Court invalidated a law forbidding bakeries to hire bakers to work for more than ten hours a day illustrates the risks inherent in judicial review. See *Lochner v New York*, 198 U.S. 45 (1905).

[28] Joseph Raz emphasised the need to distinguish clearly between the rule of law and other values of a good legal system. In his words, it is important to distinguish between the rule of law and the rule of *good* law. See Joseph Raz, "The Rule of Law and Its Virtue" in Joseph Raz, *The Authority of Law: Essays on Law and Morality* (Oxford: Clarendon Press, 1979), p. 210.

[29] Ibid., p. 221.

[30] Ibid., p. 220.

[31] *Kach Faction v Knesset Chair* 39 (3) PD 141 at 152 (1985).

dent argument which shows why the Court should have constitutional powers. Hence, this argument implicitly presupposes what it aims at proving, namely that the courts have the power of judicial review.

Moreover, even if one presupposes that the primary task of judges is to interpret the law, they can fulfil their interpretative role by declaring that the statute violates the law and yet not invalidate the statute and consequently leave the question of whether to invalidate it to another institution. The power to interpret does not necessarily entail the power of judicial review.[32]

There are, however, more fundamental controversies which lurk behind the debate over judicial review. I shall identify two primary ways to challenge the Court's argument: the one challenging the claim that the Court is indeed better at identifying the meaning of the basic laws and the second demonstrating that even if the Court identifies better than the legislature what the basic laws require, it may still violate the rule of law by invalidating statutes violating the basic laws.

The competence of the Israeli Court to interpret the basic laws cannot be examined independently of one's view of the Court's actual past performance in interpreting these laws. The opponents of judicial review in Israel often believe not only that the Israeli Court does not have the power of judicial review, but also that the Court consistently errs in its own understanding of the basic laws. The primary accusation of the opponents of judicial review is that the Court interprets Basic Law: Human Dignity and Liberty in a very expansive manner to include the right to equality and other rights which, in their view, are not included within the scope of these laws.[33] The dubious record of the Court in interpreting the basic laws inevitably leads the Court's opponents to be sceptical as to the Court's ability to interpret the basic laws better than the legislature.

Thus, beneath the dispute over judicial review, there are fundamental disagreements as to the nature of the constraints imposed on the legislature by the basic laws. The Court perceives those constraints as broad in scope, while many of its opponents perceive them to be limited. The institutional decision concerning the desirability of judicial review cannot be detached in the Israeli context from the substantial dispute over the content of the constitutional constraints imposed by the basic laws.

The examination of the greater risk premise so far was based on an examination of the delicate balance between the risk that the legislature violates its constitutional duties and legislates statutes which it does not have the authority to legislate and the risk that the Court violates its own duty by invalidating statutes which do not conflict with the basic laws. It was presupposed that if the former

[32] I am grateful to Ruth Gavison for drawing my attention to this possibility.

[33] For a survey of the different views on the question what rights are included in the term dignity, see Sommer, n. 10 above, at 257. Among the rights allegedly protected by the term "dignity", in Basic Law: Human Dignity and Liberty, Sommer includes the right to equality, the right to free speech, the right to freedom of movement, freedom of religion, freedom of association, freedom of contract, the freedom to strike and a few others.

risk is greater than the latter risk, then judicial review is conducive to the protection of the rule of law. It is time to challenge this presupposition and illustrate that even if the Court is better at interpreting the basic laws, judicial review may still be detrimental to the rule of law.

Even when a human right is protected by a basic law, the rule of law requires protecting it in accordance with the ways determined by law. Thus, the protection of rights by institutions which are not authorised by law to protect them is a violation of the rule of law even if these institutions are more competent to identify what the law is. The rule of law does not simply require maximising the instances in which the law is respected; instead it requires that the law be protected by the institutions authorised by law do so and in the ways determined by the law.[34] The Court may make decisions which are required by law, and yet violate the rule of law if the law does not permit the Court to make these decisions, or if the law does not permit the Court to make these decisions on the grounds provided by the Court. Thus the infringement of the rule of law need not rely on the prediction that the Court is likely to misinterpret the basic laws. It may instead rely on what can be labeled "institutional impropriety", namely on the claim that the Court is not the institution which has the power to make certain decisions even if it is the most competent to make them.

The impropriety of courts could of course be argued for on the basis of the general arguments against judicial review, in particular, on the basis of the claim that judicial review is detrimental to democratic rule. There are however arguments which are more specific to the Israeli legal system. Let me briefly mention two of these arguments.

First, it could be argued that although the Court should in principle be authorised to invalidate statutes, the Court should be granted this power explicitly by the legislature after an extensive public discourse. Such a discourse arguably did not precede the enactment of the two basic laws in Israel. Secondly, the Court in Israel is perceived by some as insufficiently representative of the complexity of views and ideologies and consequently granting the power of judicial review to the courts may be perceived as the political victory of a secular ideology over a religious or nationalist one. Both of these arguments do not rely on the special competence of the legislature, or the incompetence of courts. Instead, they aim to establish that even if the Court is the most competent to make decisions with respect to the basic laws, it is institutionally improper for the Court to make these decisions under the present circumstances. Moreover, under this argument, by the very attempt to take upon itself these responsibilities despite its institutional impropriety, the Court violates the rule of law even if it understands and implements the basic laws better than the legislature.

[34] This claim is analogous to the one made by Nozick with respect to rights. Under his view, respecting rights differs from minimising the total (weighted) amounts of violation of rights. One can violate a right even when she succeeds in minimising the amount of violations of rights. See Robert Nozick, *Anarchy, State and Utopia* (New York: Basic Books, 1974), p. 28.

It is not my aim to establish the claim that judicial review is undesirable; nor to establish the claim that it is undesirable in the Israeli context. The arguments in this section merely show that judicial review does not necessarily promote the rule of law. Hence, supporters of judicial review on other grounds, e.g., on the grounds that judicial review is conducive to the protection of human rights could argue that the fact that the Court uses erroneous arguments is irrelevant as long as these arguments are used to justify an important and desirable institutional mechanism, namely the mechanism of judicial review. The next section will challenge this view and demonstrate that the rhetoric of the rule of law used by courts is not only based on false premises, but may also be detrimental to lively democratic political discourse. Hence, the use of the term the rule of law to justify judicial review is not simply theoretically unsound, but also politically undesirable.

2. The rule of law and the discourse of democratic politics

Concepts and values used in judicial reasoning can be divided into different types. Some concepts are technical legal concepts constructed primarily or exclusively by legal experts, while others are primarily political concepts constructed by political activists and citizens. Concepts such as standing fall clearly into the former category—the category of technical legalistic concepts, while the precise understanding of constitutional rights and their scope is an issue which is largely determined by political discourse conducted outside of courts.

The dichotomy between these two categories is not a sharp one. Some concepts used in legal reasoning are both constructed by lawyers and by lay persons. Intermediate concepts are ones in which legal experts have a privileged, but not exclusive role in constructing their precise meaning. The rule of law is, in my view, an intermediate concept. On the one hand, it is not merely a legal term constructed by the legal community (such as standing). On the other hand, justifying a decision on the grounds that it promotes the rule of law automatically grants some privileged role to legal experts. This phenomenon was described as follows:

> "The citizen's role as distant spectators is exacerbated by the arcane and stylized language of constitutional litigation . . . In this way, legal discourse enforces its own canons of relevance, rationality and reasonableness. . . The court's historical function has not been to express popular justice, but rather 'to ensnare it, control it and to strangle it, by re-inscribing it within institutions which are typical of a state apparatus' ".[35]

The view was expressed more powerfully as follows:

[35] Allan C. Hutchinson and Patrick Monahan, "Democracy and the Rule of Law" in Allan C. Hutchinson and Patrick Monahan (eds.), *The Rule of Law: Ideal or Ideology* (Toronto: Carswell, 1987), pp. 97, 117.

> "Values such as justice and equality are the products of politics, not its antecedents. They take root in a public that engages in debate and argument that is given the opportunity to nurture notions of reasonableness and commonality. Deprived of such empowerment, public values corrode and civic energy dissipates. Deferring to 'specialists', citizens lose the capacity to define their own values and traditions. Public morality will atrophy rather than be energized. The appointment of the judicial philosopher king exacerbates the problem it was intended to remedy".[36]

These quotations point out the risks inherent in the transformation of questions of political morality into technical legal questions.[37] Such a transformation may weaken the intensity of public discourse by granting too much authority and power to specialised legal experts—"judicial philosopher kings". In deciding to justify judicial review on the grounds that it promotes the rule of law, judges transform a major political debate into the realm of specialised legal one.[38]

Arguably transforming the controversy from the discourse of politics into the discourse of law is not necessarily undesirable. Sometimes such a transformation could be necessary in order to mitigate otherwise unresolvable disagreements. Stephen Holmes, for instance, argues persuasively that the American system has resolved some of the most divisive issues in society such as the debates over slavery and abortion by turning these issues over to courts and thereby freeing public discourse to engage in issues which can be resolved peacefully by the political process.[39] More specifically, Holmes believes that:

> "By the narrowing of the political agenda to problems manageable by discussion, certain individual rights may be said to subserve self-government. Their function, once again, is not merely to shield the private but also to disencumber the public".[40]

This is an important observation. But the viability of such a solution depends upon certain presuppositions. Most importantly, it depends upon a prior public consensus concerning the powers of the courts and trust that these powers be used to promote the interests of all. The relative success of the American system

[36] Ibid., p. 119.

[37] In fact the Court itself is not oblivious to these risks. In *Ressler*, Chief Justice Shamgar said as follows: "There are cases where consideration of a particular issue according to legal standards alone will miss the point because it is likely to obscure the true nature of the problem under consideration. Frequently it is not the legal norm which gives rise to the problem and it has no decisive significance for the substance of the decision": see Ressler (the English translation), n. 5 above, at 124–5.

[38] It may seem that my analysis presupposes the Marxist view that legal concepts are often abused for the sake of legitimating elites' power. See Michael Mandel, "Marxism and the Rule of Law" (1986) 35 *University of New Brunswick L.J.* 7. See also the discussion in Hugh Collins, *Marxism and the Law* (Oxford: Oxford University Press, 1986), ch. 6. In contrast, however to the Marxist analysis, my analysis is limited to Israeli courts and I have no inclination to argue that this phenomenon is universal. Israeli courts operate under very difficult institutional pressures and it is difficult to extend from the Israeli experience to the experience of other courts.

[39] See Stephen Holmes, "Gag Rules or the Politics of Omission" in Jon Elster and Rune Slagstad (eds.), *Constitutionalism and Democracy* (New York: Cambridge University Press, 1988), p. 19.

[40] See ibid., p. 24.

in eliminating some of the most divisive issues from the political discourse was premised upon a certain consensus concerning the powers of the courts; or at least the absence of great resistance or suspicion of the courts. Courts rely on this consensus when they remove an issue like abortion from the political discourse and resolve it by transforming it into a legal question. Courts cannot, however, effectively resolve the very controversy concerning their own powers to invalidate laws when such an issue has become politically controversial.

By claiming that the term the rule of law functions as a mechanism to transform the nature of the discourse, I do not claim that courts use this term deliberately in order to silence their opponents. Judicial reasoning is part of a ritualised discourse which courts conduct with other segments of the Israeli society. This discourse is not meant to convert the opponents, or to silence them, but it conveys the feelings that courts are somehow senior, or privileged participants in the public discourse over the Israeli Constitution.

The argument that judicial review promotes the rule of law does not merely entail that the Court should have the power of judicial review. It also implies that the Court can understand and judge better than its opponents why judicial review is so important. Under this understanding, judges are, by virtue of their legal expertise, the most senior participants in the constitutional controversy concerning judicial review. Thus, Israeli courts use the term rule of law in a manner which transforms a legitimately political debate over the powers of the Court into a legalistic specialised debate in which legal experts have a privileged role. The use of the term the rule of law in this context transforms the heated legitimate political debate over the powers of the Court into a specialised legalistic debate—a debate in which judges have a special expertise. The attempt of the Court to describe judicial review as an essential component of the rule of law is in my view an indication that the institutional needs of the Court, in particular, its effort to expand its constitutional powers, influence its understanding of the concept of the rule of law.

IV. CONCLUSION

Fuller rightly identified the indeterminacy of the concept of the rule of law, an indeterminacy which he described by stating that the rule of law is grounded in "the morality of aspiration".[41] His insight shows that courts have an important role in shaping the precise meaning of this term. Courts, however, may use the term in a way which is conducive to their institutional needs. The use made of the term "the rule of law" is always a byproduct of a complex interaction between the theoretical understanding of the term and institutional needs of the Court.

It would be too cynical and in my view false to presuppose that courts, or other institutions, always manipulate the concept of the rule of law in a way

[41] See text to n. 1 above.

conducive to their institutional interests. Courts do not have monopoly over the meanings of terms such as the rule of law and they are constrained by the societal understandings of the term the rule of law as well as by their own ideological commitments.

But while manipulation is not inevitable, this essay illustrates that in the Israeli context, the term "the rule of law" operates in a way which serves the institutional needs of courts. The recent, opposition to the expansion of the powers of the Court as well as the relatively open-ended meaning of the term "the rule of law" leads the Court inevitably to use the term in a way which promotes its own understanding of its role in Israeli society. Moreover, by claiming that judicial review is conducive to the rule of law, judges do not merely expand their powers; they also fortify a privileged position within constitutional law discourse.

Judicial review may be an important tool in protecting human rights. Its desirability however does not entail that judicial review is conducive to the rule of law. The reasoning used by judges to justify their constitutional understandings should be politically persuasive and yet it should also be conceptually sound. The claim that judicial review promotes the rule of law is a powerful and persuasive rhetorical tool and yet it fails to withstand rigorous scrutiny.

Part II

Reconceiving the Rule of Law

8

Rhetoric and the Rule of Law

NEIL MACCORMICK*

INTRODUCTION

A concern for the rule of law is one mark of a civilised society. The independence and dignity of each citizen is predicated on the existence of a "governance of laws, not men". Values like legal certainty and legal security are much lauded in the context of praising the rule of law. The idea that law exhibits an inner morality in the achievement of these, always imperfectly but not always with equal imperfection, is vigorously canvassed. And yet the law which rules in a civilised society seems to be a site for never-ending argument and dispute. Almost everything to do with law can apparently be problematised in one context or another, and arguments adduced for doubting the most cherished certainties. As ancient as any appeal to the rule of law is recognition of law's domain as a locus of argumentation, a nursery of rhetoric in all its dubious arts. And rhetoric can here turn in on itself. Argument from commonplaces or truisms is common in rhetoric, but the commonplace truths of everyday thinking may sometimes seem at least to be in flat mutual contradiction. The idea of the arguable character of law seems to pour cold water on any idea of legal certainty or security. Or is it so? Is there any prospect of reconciling these two? What follows will apply some reflections on rhetoric and argumentation in law with a view to such a reconciliation.

I. FIRST COMMONPLACE: THE ARGUABLE CHARACTER OF LAW

Law is an argumentative discipline. Whatever question or problem one thinks about, if we pose it as a legal question or problem, we seek a solution or answer in terms of a proposition that seems sound or at least arguable, though preferably conclusive, as a matter of law. To check whether it is sound or genuinely arguable, or perhaps even conclusive, we think through the arguments that could be made for the proposed answer or solution. We can test such arguments

* This essay is a modified version of one presented to the Perelman Symposium at the 20th World Congress in Philosophy, held in Boston in August 1998, and to be published in the printed proceedings of the Congress.

by constructing all the counter-arguments we can think of. If this be said on one side of the argument, that will be said on the other side. By thinking out what seems to be the strongest argument or strongest arguments on that side, we test the strength of the arguments on this side. By figuring out the counter-case they have to meet and, if possible, defeat, lawyers get their arguments into the best shape possible.

This is not an exact science, for it is not a science at all but a practical skill, a practical art. Yet it very much depends upon knowledge and learning (law is not inaccurately called a "learned profession"). Legal arguments are always in some way arguments about the law, or arguments about matters of fact, of evidence, or of opinion, as these have a bearing upon the law, or as the law has a bearing on them. To know, and indeed to be intimately familiar with, a great body of legal learning, is essential both to the making, and also to the evaluating, of high quality arguments in law. So legal science, the structured and ordered study of legal doctrine, is one essential underpinning of law as praxis. Many persons of deep learning evince little flair for forensic argumentation; some persons of considerable flair and skill lack the application fully to master the law. It is the combination that is required.

A process of evaluating the relative strength of competing arguments is bound to be a matter of more-or-less, a matter of opinion, calling for judgement. If arguments often seem close-matched, how can we tell for certain which is the stronger? Probably the answer is that we can't say with certainty, not as we can in demonstrative arguments,[1] where acceptance of premises as axiomatic or as contingently true allows us to derive from them a conclusion which cannot be doubted so long as its premises stand. In law, subjective conviction is possible on occasion, where for you or for me a certain body of arguments points firmly to a certain conclusion, and all the counter-arguments that have been put to us or that we can think of seem fatally weak by comparison. And this can be a shared or inter-subjective certainty, when a community of experts[2] shares such a view, to the point even of treating it as practically axiomatic. But such a shared conviction, such a shared attitude of being certain about something, is not what is meant by certainty in the other sense: that which is certainly true, whether anyone actually believes it or not.

All this is, I suppose, relatively commonplace among those who have any interest in law, whether as a subject of study or as a practical profession. It is the kind of common opinion which leads on into related positions such as: that the law is not logical; that logic contributes nothing to legal argument; that law has nothing to do with truth, only with what can be proved according to the law's

[1] Ronald Dworkin is a particularly forceful proponent of the well-taken point that the non-demonstrative character of legal arguments is not a bar to their being nevertheless sound arguments, in a context in which one sound argument can genuinely defeat another. See *Law's Empire* (London: Fontana Books, 1986), pp. 9–15, and compare S. Guest, *Ronald Dworkin* (Edinburgh: Edinburgh University Press, 1991), pp. 141–4.

[2] Compare Aulis Aarnio, *The Rational as Reasonable* (Dordrecht: Kluwer Academic, 1987), pp. 221–5.

processes and standards of proof, applied to whatever evidence the law characterises as relevant and admissible. Whether such derived positions really are necessary corollaries of our commonplace starting point is far from obvious, and will be disputed in this work. But the starting point itself will stand as one key element in what must be grasped by one who would understand the nature and character of law as practical activity.

II. SECOND COMMONPLACE: THE RULE OF LAW

This, then, we take as an opening thesis about law: so far as law is that which underlies legal claims or accusations and legal defences, law is something arguable, sometimes, but not always, conclusively, always at least persuasively. To this must then at once be posed an antithesis, that also belongs to the merely commonplace: where law is faithfully observed, the rule of law obtains; and societies that live under the rule of law enjoy great benefits by comparison with those that do not. The rule of law is a possible condition to be achieved under human governments. Among the values that it can secure, none is more important than legal certainty, except perhaps its stable-mate, security of legal expectations and safety of the citizen from arbitrary interference by governments and their agents.[3]

Where the rule of law is observed, people can have reasonable certainty in advance concerning the rules and standards by which their conduct will be judged, and the requirements they must satisfy to give legal validity to their transactions. They can then have reasonable security in their expectations of the conduct of others, and in particular of those holding official positions under law. They can challenge governmental actions that affect their interest by demanding a clear legal warrant for official action, or nullification of unwarrantable acts through review by an independent judiciary. This is possible, it is often said, provided there is a legal system composed principally of quite clearly enunciated rules that normally operate only in a prospective manner, that are expressed in terms of general categories, not particular, indexical, commands to individuals or small groups singled out for special attention. The rules should set realistically achievable requirements for conduct, and should form overall some coherent pattern, not a chaos of arbitrarily conflicting demands.[4]

Many people, and certainly I for one, find attractive both thesis and antithesis as stated above. I do believe in the argumentative quality of law, and find it admirable in an open society. We should look at every side of every important question, not come down at once on the side of prejudice or apparent certainty. We must listen to every argument, and celebrate, not deplore, the arguable

[3] Compare J. Raz, "The Rule of Law and its Virtue", in Raz, *The Authority of Law* (Oxford: Clarendon Press, 1979), pp. 210–29.

[4] The *locus classicus* for this type of account remains L. L. Fuller, *The Morality of Law* (New Haven, Conn: Yale University Press, revised edn., 1969), ch. 2.

quality that seems built in to law. But I also believe in the rule of law, and think that our life as humans in community with others is greatly enriched by it. Without it, there is no prospect of realising the dignity of human beings as independent though interdependent participants in public and private activities in a society. Dignity of that sort and independence-in-interdependence are, to my way of thinking, fundamental moral and human values.

How is it possible to believe in both? Can this be anything other than wishful believing? These are the questions that lie before us. Can we reconcile the commonplace of the 'Arguable Character of Law' with the ideology of the 'Rule of Law'?

III. TOWARD RECONCILIATION

1. Rhetorical theories

The strategy of trying to reconcile competing commonplaces that I shall adopt here depends on acknowledging a fundamental constraint on the process of legal argumentation. This lies in the so-called 'special case thesis' suggested by Robert Alexy.[5] That is to say, legal argumentation must be acknowledged to be one special case of general practical reasoning, and must thus conform to conditions of rationality and reasonableness that apply to all sorts of practical reasoning. This implies at least that there may not be assertions without reasons—whatever is asserted may be challenged, and, upon challenge, a reason must be offered for whatever is asserted, whether the assertion is of some normative claim or a claim about some state of affairs, some "matter of fact".

Thus it is a restricted version of the Arguable Character of Law that will be reviewed and defended here. The argument will be confined to considering what is rationally arguable. To say this is to distinguish between the use of words as mere weapons of intellectual coercion or deceit, and their use as instruments of reasonable persuasion, where coercion appears only in the sense of the compelling force of an argument. It is the latter, argument as rational justification, that will be reviewed here. And the issue will be whether there can be a "Rule of Law", if "law" is a matter of what is arguable in this sense. It will remain an open empirical question whether or how far actual attorneys and judges in any particular state confine their use of argumentation to the domain of the practically reasonable.

Notwithstanding the restriction to what is rationally arguable, the very idea of law as arguable leads us at once to consider the rhetorical character of legal

[5] See R. Alexy, *Theory of Legal Argumentation* (R. Adler and N. MacCormick (trans.), Oxford: Clarendon Press 1989), pp. 5–10, 212–20, 294–5; and compare N. MacCormick, *Legal Reasoning and Legal Theory*, 2nd edn. (Oxford: Clarendon Press, 1978, 1994), making substantially the same point (pp. 272–4); Alexy's *Theorie der juristischen Argumentation* was first published in 1978, almost exactly contemporaneously with *Legal Reasoning and Legal Theory*, so we each came to much the same view by independent processes of discovery.

argumentation. Wherever there is a process of public argumentation, there is rhetoric. The modern rediscovery of rhetoric as a discipline owes much to reflection on legal reasoning. Theodor Viehweg, drawing on Aristotle, has drawn attention to the significance of *topoi*, or "commonplaces" in rhetorical arguments.[6] An argument for a particular rule or proposition can be supported by reference to some accepted *topos*, and arguments progress by working towards, or from, such commonplace positions. In law, there are maxims and long-standing principles and presumptions, such as "a person is to be presumed innocent until proven guilty", "no one can give a better right than he/she has himself/herself" or "a later law derogates from an earlier one" and such like. Likewise, there are well-established argument forms such as *argumentum a fortiori, argumentum a maiori ad minus, argumentum per analogiam* and the like. An argument in such a recognised form starting from or working toward a recognised *topos* is well-calculated to be persuasive in its given context. The present argument itself starts from two ideas that seem to me well-established commonplaces among those who think, even sporadically, about law.

Using a concept not far removed from that of "commonplace", Duncan Kennedy suggests that common law arguments typically proceed through the adduction of standard "argument bites", of a kind that frequently can be found in matching pairs, so that a persuasive legal argument will be an aggregation of argument bites relativised to the fact-situation in question, but a counter-argument can be constructed using a similarly contextualised set of counter-arguments in the form of matching "bites".[7] James Palmer has shown that this insight may be exploited in harnessing information technology and artificial intelligence to assist in processes of legal reasoning. Intelligent knowledge-based systems can be envisaged that would generate a battery of relevant argument bites for adduction in relation to problems in given domains of law. So far, at least, there is no suggestion that the evaluation of competing arguments constructed in this way could or should be delegated to computers; rather, the hope would be to ensure that lawyers and judges would come to the task of constructing their final arguments to lay before a court, or to deploy in justification of a decision, with a thoroughly worked-over checklist of available arguments based on prior practice (precedent) and, where appropriate, statute law.[8] The present argument as posed so far itself takes two bites or *topoi* and sets them in mutual opposition. The question is where we can get from this apparent contradiction.

[6] See T. Viehweg, *Topik und Jurisprudenz*, 5th edn. (Munich: C.H. Beck, 1974); also "Some Considerations concerning Legal Reasoning", in G. Hughes (ed.) *Law, Reason and Justice: Essays in Legal Philosophy* (New York: New York University Press, 1969), pp. 257–69 at pp. 266–8; here, "*topoi*" is translated as "points of view"; see also Alexy, *Argumentation*, pp. 21–4.

[7] See D. Kennedy, *A Critique of Adjudication* (Cambridge Ma: Harvard University Press, 1997), pp. 137–56, and note other "CLS" authors and works cited at ibid., p. 393; as Kennedy fully acknowledges, the fundamental insight here goes back at least as far as to the early work of Karl Llewellyn—see, e.g., Llewellyn, *Jurisprudence: Realism in Theory and Practice* (Chicago: Chicago University Press, 1962), pp. 70–1).

[8] See J. Palmer, *Artificial Intelligence and Legal Merit Arguments* (Oxford: Oxford University D. Phil. Thesis, 1998), pp. 109–28.

Certainly, one should bear in mind Josef Esser's teachings concerning the importance of *Vorverständnisse,* "pre-understandings", the taken for granted assumptions that enter any judgement of what is acceptable in the setting of legal argumentation—and in the preference of one method of arguing over another in a particular case. Once premises and mode of argument are settled, it is relatively easy to produce an argument that satisfactorily justifies the conclusion reached. But the problem then becomes one about the reasonable choice of premises and method, so there must be inquiry into pre-understandings.[9] Aulis Aarnio has suggested that in the end these may simply have to be assessed as the "form of life" that they constitute.[10]

La Nouvelle Rhétorique of Chaim Perelman emphasises that arguments are necessarily addressed to an audience, and that persuasiveness is audience-relative. This is specially obvious in legal practice, where trained advocates put cases before courts as persuasively as possible, and judges decide after weighing their rival arguments on points of law. In systems where juries are responsible for the determination of facts, or of legal conclusions reached through their own findings of fact in the light of law as explained by the presiding judge, the rhetorical character of forensic argumentation is yet more salient. But from the point of view of practical rationality, the immediate and concrete persuasiveness of an argument is not necessarily the same as its soundness. The issue for a theory of reasoning-as-justification is not what argument actually persuades a particular judge or jury, but what ought to convince any rational decision-maker. In this connection, Perelman postulates a "universal audience" as providing the ultimate test: whatever argument would convince the audience of all intelligent and concerned persons, evaluating issues in a disinterested way, is a sound one.[11]

More or less contemporary with Perelman's work on rhetoric was Stephen Toulmin's on *The Uses of Argument*. This offered a way of narrowing the apparent gap between the supposedly timeless pure rationality of formal logic and the context-bound character of rhetorical argumentation and persuasion. Toulmin proposed a re-interpretation of traditional logic as a sort of normative ordering of thought processes and public presentations of reasons, a process that regulates moves in the play of arguments. Rather as a ticket entitles one to undertake a certain journey by train or plane, appropriate forms of argument supply warrants that entitle one to move from premises to conclusions. The validity of the move depends on the soundness to its context of the warrant produced.[12]

[9] J. Esser, *Vorverständnis und Methodenwahl in der Rechtsfindung* (Frankfurt/Main: Suhrkamp, 1970), pp. 3–20.

[10] A. Aarnio, *Rational as Reasonable*, n. 2 above, pp. 213–18.

[11] Ch. Perelman and L. Olbrechts-Tyteca, *La Nouvelle Rhétorique* (Paris: Presses Universitaires de France, 1958); *The New Rhetoric* (J Wilkinson and P. Weaver (trans.) Notre Dame, Ind: Notre Dame University Press, 1969), see pp. 76–86 on the "universal audience".

[12] S. E. Toulmin, *The Uses of Argument* (Cambridge: Cambridge University Press, 1958), pp. 94–145.

The rhetorical turn in analysis of practical reasoning is unserviceable to the present purpose if (or in so far as) it reduces the rational acceptability of an argument to its actual persuasiveness. One of the things that gets rhetoric a bad name is the notorious possibility that a good speaker can win an audience round with a bad case. To counter this with appeals either to the universal audience or to some supposed consensus of reasonable contemporaries seems question-begging, since in fact we work out what would persuade the universal audience by reference to what is sound, not vice versa, and we have no guarantee that a contemporary consensus, where it exists, is correct. Yet again, the "critical" approaches to legal thought urge that the claim to an objective soundness of legal reasons is the grandest rhetorical turn of all.

This point of the "Critics" has in part been noted already. Often a set of persuasive reasons or "argument bites" can be built up to give strong support for one solution to a legal problem or controversy. But in any actually or imaginedly controversial situation we can find a matching counter-reason or counter-bite for each of them. So the problem is not to uphold a soundly arguable case at the expense of a manifestly weaker case. Rather, it is all too often a matter of choosing between two strongly arguable and strongly argued cases, in a dialectical situation in which each argument made by either party is firmly countered by a good argument proposed by the other. Perhaps, therefore, it is only by reference to considerations of ideology extraneous to law that one can come to a justified decision at all, and the ultimately justifying ground is ideology, not law.[13] Hans Kelsen's brief discussion of interpretation in *The Pure Theory of Law* points in the same direction as this.[14]

2. Proceduralist theories

A procedural approach to practical reasoning may, however, provide a partial solution to the problems posed concerning rhetoric. There are various "proceduralist" approaches, but they have in common a concern with understanding the constraints on practical reasoning that have to be acknowledged if it is to yield rationally acceptable conclusions in an interpersonal context. So a starting point is indeed the rhetoricians' emphasis on the interpersonal context of argumentation. In its light, the concept of universality has two uses. First, it demands universalisability of reasons—for the present instance of circumstances C to count as a reason now for reaching decision D, and acting on D, it would have

[13] See J. Balkin, "The Crystalline Structure of Legal Thought" (1986) 39 *Rutgers Law Rev* 195; "Nested Oppositions" (1990) 99 *Yale Law Journal* 1669; "Ideological Drift and the Struggle over Meaning" (1993) 25 *Conn. Law Rev* 369; compare Kennedy, *Adjudication*, n. 7 above, pp. 133–8, and see Peter Goodrich, *Reading the Law: a Critical Introduction to Legal Method and Techniques* (Oxford: Blackwell, 1986), pp. 213–23.

[14] See H. Kelsen, *The Pure Theory of Law* (M. Knight (trans.) Berkeley and Los Angeles: University of California Press, 1967), pp. 251–4 on the "political" character of interpretative decisions taken within the framework of statute law.

to be acceptable to hold a decision of type D appropriate whenever an instance of C occurs. Secondly, it suggests a way of testing whether it is warranted to assert that D is appropriate whenever C obtains. This universalised reason, by its terms, will be applicable to all instances of C, not just the single instance now under attention. The interests, feelings, and opinions of all human beings are therefore potentially at stake, and one can ask whether the formula "Whenever C, then D" could be rejected by anyone who is willing for everyone to have the same opportunity to challenge practical principles of decision.[15]

As Jürgen Habermas and followers like Robert Alexy argue, it may be possible to test practical propositions by reference, at least in principle, to the interests, feelings, and views of the totality of persons in any way affected by or concerned with them. Habermas's move is to propose a test by reference to dialogue in an "ideal speech situation", envisaged as one in which all forms of coercion or interpersonal power or domination are put aside for the purposes of conducting (or imagining the conduct of) interpersonal discourse. Analysis of the necessary constraints on such a discourse yields a procedural approach to testing the kinds of principles that rational discourse-partners could accept, acknowledging the types of desires and interests they actually have.[16] Important in this is the idea that accepted principles or commonplaces (*topoi*) should be subject to challenge, but are considered acceptable until successfully challenged, e.g., on the ground that they cannot pass the test of universalisability or on the ground that they owe their origins to past or present social power-relations that would themselves have been rejected in the ideal speech situation.[17] A similar, but simpler and thus more persuasive, idea is that of T. M. Scanlon,[18] who suggests that an action is wrong if any principle that permitted it would be one that, for that reason, someone could reasonably reject even if that person were moved to find principles for the general regulation of behaviour that others, similarly motivated, also could not reasonably reject.[19]

It is doubtful whether any such procedural approach wholly disposes of recourse to personal feelings or subjective intuitions. For one has to interrogate the grounds that would lead one to reject a certain principle oneself, and still the

[15] Compare Alexy, *Argumentation*, n. 5 above, pp. 65–9, 146–7, 267–77; MacCormick, *Legal Reasoning*, n. 5 above, pp. 76–86.

[16] Habermas, *The Legitimation Crisis* (A. McCarthy (trans.), London: Heinemann, 1976), pp. 109–12; compare Alexy, *Argumentation*, n. 5 aabove, pp. 111–37.

[17] See Alexy, *Argumentation*, n. 5 above, pp. 151–3, 204–5.

[18] See T. M. Scanlon, "Contractualism and Utilitarianism", in A. Sen and B. Williams (eds.), *Utilitarianism and Beyond* (Cambridge: Cambridge University Press, 1982), pp. 103–28; and *What We Owe to Each Other* (Cambridge, Ma: Harvard University Press, 1998). Since no actual agreement or contract is involved in such reasoning, and obligations generated by or under it are not in fact contractual in character, it seems to me regrettable that this style of procedural testing of practical principles has been dubbed "contractualist"; but its value as a mode of reasoning is unaffected by the name it bears. Compare MacCormick, "Justice as Impartiality: Assenting with Anti-contractualist Reservations" (1996) *Political Studies* 305–10.

[19] Quoted from T. M. Scanlon, "Promises and Contracts", preliminary draft of a paper prepared for P. Benson (ed.), *Philosophy and Contract Law* (Cambridge, Cambridge University Press, forthcoming).

question has to be faced what it is "reasonable" for anyone to reject given the feelings and pre-understandings that each person brings to the judgement-seat. The procedure of procedurally testing arguments cannot be infinitely regressive. But it is surely a merit of such procedural approaches that they both postpone and narrow appeals to intuition[20] and to gut-feeling. They enable us to scrutinise claims about what is reasonable in the light of acknowledged constraints of rational discourse. Commonplace principles are still needed as starting points, but they are challengeable within the argumentation.

Rationality of argumentation introduces another significant constraint. Although any particular practical dilemma or topic of concern falls to be considered on its own merits, and subjected to procedures such as we have considered so far, one must recall that the universals ("whenever C, then D") that we work toward cannot be envisaged as once-off isolated commitments. We who decide them do so as part of an ongoing and interpersonally engaged social life in which decisions and dilemmas are recurrent in character. This has a strong bearing on what one can reasonably accept or reject in terms of the Scanlonian meta-principle or the Habermasian ideal speech situation. So one's principles and rules of decision and of conduct have to belong in a body of practical thought and commitment that is internally consistent, and characterised also by a certain overall coherence.[21] This implies at least some guidelines about priority-rankings and procedures to determine relative weights of practical reasons in order to resolve *prima facie* conflicts.

Here, it is useful to remind oneself of the starting point of the present train of reasoning, namely in a puzzle about the apparent antinomy between law as that which is arguable, and law as that which guarantees security and stability in social life within a *Rechtsstaat*. So far, we have considered in a somewhat abstract way the idea of a rhetorical engagement in the practice of the law, and how far a procedural or discourse-theoretical development of ideas from the "new rhetoric" offer the hope of an acceptable rational framework for our argumentation in law and indeed in other practical domains.

The legal context, however, is one in which the recently mentioned idea of coherence has a particular and obvious significance.[22] In a legal argument, no one starts with a blank sheet and tries to work out a reasonable conclusion *a priori*. A solution offered must ground itself in some proposition that can be at least colourably presented as a proposition of law, and such a proposition must be shown to cohere in some way with law as already determined. Legal

[20] M. J. Detmold, *The Unity of Law and Morality* (London: Routledge, 1984), at pp. 115–22 rightly stresses the element of intuition in judgment, but leaves too wide open the scope for its exercise.

[21] See A. Peczenik, *On Law and Reason* (Dordrecht/Boston/London: Kluwer Academic Publishing, 1989).

[22] See Kennedy, *Adjudication*, n. 7 above, pp. 33–4; N. MacCormick, "Coherence in Legal Reasoning", in W. Krawietz et al. (eds.) *Theorie der Normen* (Berlin: Duncker & Humblot, 1984), pp. 37–53 and "Time, Narratives, and Law" in J. Bjarup and M. Blegvad (eds.), *Time, Law, and Society* (Archiv für Rechts- und Sozialphilosophie Beiheft 64), pp. 111–125.

argument-makers and decision-makers do not approach problems of decision and justification in a vacuum, but rather in the context of a plethora of material that serves to guide and to justify decisions, and to restrict the range within which the decisions of public agencies can legitimately be made.

The material in question comprises constitutions, treaties, statutes of national parliaments, regulations and directives of supranational entities, and the multitudinous reports of decisions by judicial tribunals, recognised in some systems as "precedents" in the sense of a "formal source of law", and used in practically all systems of law[23] as at least a repertory of available guides to interpretation of statutes, constitutional articles, and other formally binding legal provisions. It also includes treatises and other scholarly writings on law by acknowledged legal experts.

3. Laws

It is trite to say that this mass of material is not and cannot be imagined to be self-interpreting and self-applying. In the perspective of the "Rule of Law" ideal, it has to be comprehended as the raw material of a "legal system", organised in intelligible bodies of material relevant to particular human concerns within traditionally understood branches or domains of law, such as property, contract, family law, criminal law, administrative law.[24] In the context of states as coercive associations of human beings, governments and thus the human beings who perform governmental roles are empowered to act authoritatively towards others, and can back their assertions of authority with decisions to deploy organised coercive power. Here, the demand for rational justifiability of governmental action is an urgent one if government is not to be the mere mask of tyranny. Hence it has come to be generally understood as legitimate to demand that any governmental act be warranted by explicit provisions mandating, permitting or authorising decisions in specific terms (or involving some bounded discretion) only when certain quite clearly specified circumstances obtain. Provisions of this kind, especially when specifically enacted by some legislative process, but also when they can be derived in reasonably definite terms from other materials such as precedents, are typically called "rules", in contradistinction to other kinds of norm, such as conventions, standards, values or principles.[25]

A legal rule is a normative provision stated in or constructed from a recognised legal source that has the form of linking a determinate normative conse-

[23] See D. N. MacCormick and R. S. Summers, *Interpreting Precedents: A Comparative Study* (Aldershot: Dartmouth Publishing Co, 1997).

[24] See J. Wróblewski, *The Judicial Application of Law* (Z. Bankowski and N. MacCormick (eds.), Dordrecht/Boston/London: Kluwer Academic, 1992), pp. 75–85.

[25] For a clarification of these distinctions, see N. MacCormick, "Norms, Institutions, and Institutional Facts" (1998) 17 *Law and Philosophy* forthcoming.

quence to determinate operative facts. It is in the nature of a rule to provide that whenever a certain state of facts obtains, a given normative consequence is to follow therefrom. To put this in a standard form: "Whenever OF then NC".

At the heart of the liberal idea of free government, and at the heart of the distinction between free and despotic governments is the idea that when governments act towards citizens, their action must be warrantable under a rule in this sense of the term; and this holds good also when government, usually through the agency of the judiciary, purports to regulate or pass judgment on claims and complaints and demands levied by citizen upon citizen Here too the rule of law demands that there be some rule to warrant the claim of one person against another if adjudication of the claim is liable to issue in an enforceable order against that other, for example an award of compensatory or punitive damages or an injunction or interdict.

Codes and statutes of the modern period, and other like materials, represent an institutional response to the ideology[26] of the rule of law as a condition of liberty. The state that governs through law takes care to provide in advance the rule-texts which warrant public interventions in private lives, whether such interventions be prompted by public authorities or by private litigants. The security for individuals that is thus guaranteed consists in the fact that rule-application evidently requires the prior existence of specific facts instantiating the relevant rule's generically stated operative facts "OF". Thus, for example, if a statute provides a remedy for persons who suffer discrimination in their employment "because of sex", no action to implement the normative consequence(s) that the rule provides for can be justified unless in a particular case some act of discrimination has occurred, has occurred in the context of an employment relationship, and is attributable to the sex of the person discriminated against.[27] Or if a rule provides for nullification of a driver's licence and for some other penalty within a range determined at a judge's discretion when a person drives a motor vehicle while impaired by the consumption of alcohol beyond a specified proportion in her/his bloodstream, no penalty is legitimately exigible against a particular person except if warranted by her/his having been in the condition specified and having "driven" a "vehicle" while in that condition.[28]

If the rule of law is to be actually a protection against arbitrary intervention in people's lives, it seems clear that it is not in practice enough to demand that

[26] "Ideology" in this context is of course not used in its pejorative sense. Compare Wróblewski, *Judicial Application*, n. 24 above, pp. 75–85.

[27] See *Oncale v Sundowner Offshore Services*, US Supreme Court No 95–568 (March 1998) where the Supreme Court held that same-sex harassment could amount to a breach of the anti-discrimination provisions in Title VII of the Civil Rights Act 1964.

[28] See *State v Blowers*, 71 A.L.R. 4th 1121 (1986, Utah) on the question whether a person riding a horse while drunk can be said to be in charge of a "vehicle"; J. Palmer, *Merit Arguments*, n. 8 above, pp. 36–8, cites a fascinating run of precedents from various jurisdictions illustrating the indeterminacy of the predicate "drive" in the context of judicial determinations about the statutory offence committed by those who "drive" a motor vehicle while disqualified from doing so.

the operative facts did on some occasion actually happen or obtain. It is necessary that some specific and challengeable accusation or averment of relevant facts be made to the individual threatened with action, and that it be supported by evidence in an open proceeding in which the party charged may contest the evidence item by item and in its cumulative effect, and may offer relevant counter-evidence as she/he chooses. Moreover, it must also be possible to challenge the relevancy of the legal accusation or claim on the ground that, whatever be the facts, the legal materials that supposedly warrant the assertion of a rule governing the case do not warrant it at all in the alleged, or the actually proven, state of the facts.

Here we are on the familiar terrain of the relative indeterminacy of law.[29] This indeterminacy is in a curious way magnified by the very same considerations that lead to the demand for determinate law. For the dialectical or argumentative character of legal proceedings is a built-in feature of a constitutional setting in which citizens are able to challenge the allegations of fact and the assertions of law on the basis of which government agencies of their own volition or at the instance of private litigants threaten to intervene coercively in their lives or affairs. A vital part of the guarantee of liberty in the governing conception of the rule of law is that the opportunity to mount such a challenge on fair terms and with adequate legal assistance be afforded to every person. And yet that same governing conception calls for relatively clear and determinate law in the form of pre-announced rules.

Hence legal indeterminacy is not merely (though it is also) a result of the fact that states communicate their legal materials in natural ("official") languages, and that these are afflicted with ambiguity, vagueness and open texture.[30] It also results from, and is in some measure magnified by, the due recognition of the "rights of the defence" in every setting of criminal prosecution or civil litigation. Every doubt that can be raised against prosecutor or plaintiff, whether concerning fact or concerning law, is a doubt that may be raised by the defence. On the other hand, wherever there seems to be a significant point of justice or of public order in issue, the plaintiff or prosecutor has reason to seek in the materials of the law some provision that will, upon some reasonably arguable interpretation, justify the civil action or criminal prosecution brought in the given case. And the defence will then again challenge what it characterises as a strained or illegitimate reading of the law according to how courts, lawyers and citizens have previously understood and acted upon it.

Thus emerge contests over proper interpretation of legal materials, over the proper drawing of inferences from evidence, over evaluation of conflicting

[29] See S. J. Burton, *An Introduction to Law and Legal Reasoning*, 2nd edn. (Boston/Toronto: Little Brown & Co, 1995), pp. 27–8, 54–8, 77–85.

[30] Compare H. L. A. Hart, *The Concept of Law*, 2nd edn. (Oxford: Clarendon Press, 1994), ch. 7, and compare B. Bix, *Law, Language, and Legal Determinacy* (Oxford: Clarendon Press, 1993), pp. 7–35, MacCormick, "On Open Texture in Law" in P. Amselek and N. MacCormick (eds.), *Controversies about Law's Ontology* (Edinburgh: Edinburgh University Press, 1991) pp. 72–84.

pieces of evidence, over the proper characterisation of facts proven or agreed, or over their relevance to the legal materials adduced.[31] These contests are not some kind of a pathological excrescence on a system that would otherwise run smoothly. They are an integral element in a legal order that is working according to the ideal of the rule of law, so far as that insists on the production by governments of an appropriate warrant in law for all that they do, coupled with the right of the individual to challenge the warrant produced by government.

This leads to an obvious conclusion. Although it may be possible to formulate rules in a verbally straightforward formula, "Whenever OF then NC", in any contested case a challenge can be raised in one or more of these ways:

(1) no instance of "OF" as alleged in indictment or pleadings has been proven (up to the required standard of proof) to have existed, taking account of all relevant and admissible evidence, including any evidence in rebuttal adduced by the defence (we may call this the "problem of proof");
(2) what has been alleged, whether or not proved, is not properly characterized as an instance of "OF" in the sense proper to the law (we may call this the "problem of characterisation" or of "classification", or of "qualification");
(3) the case as presented depends on reading the acknowledged rule "Whenever OF then NC" according to a particular interpretation of "OF" or of "NC" or both; but this is a misinterpretation, and there is in fact a more legally acceptable interpretation according to which the defence ought to be absolved from the accusation or claim laid against it (we may call this the "problem of interpretation");
(4) success in the claim or prosecution depends on reading authoritative legal materials as though they generated a rule "Whenever OF then NC" such that the allegations of criminal guilt or civil liability are relevant given the facts alleged, or even the facts proven; but no such norm can properly be read out of the adduced materials as a reasonable concretisation of them or determination from them (we may call this the "problem of relevancy").

We may now move towards a conclusion. What we see is how legal processes move through a chain of putative certainties that are at every point challengeable. No claim or accusation may be made without proper citation of the legal warrant that backs it and without giving notice of the allegations of fact in virtue of which it is asserted that the law warrants the conclusion proposed (by prosecutor or by plaintiff). This has the full logical certainty that inheres in syllogistic form.[32] There is a rule "Whenever OF then NC", cited by prosecutor or plaintiff in indictment or in pleadings, and it is there also alleged that "OF" has

[31] Compare MacCormick, *Legal Reasoning*, n. 5 above, pp. 65–72.

[32] But compare for an opposed view B. S. Jackson, *Law, Fact and Narrative Coherence* (Liverpool: Deborah Charles, 1988), pp. 37–60; in response to which, see N. MacCormick, "Notes on Narrativity and the Normative Syllogism" (1991) IV/11 *International Journal for the Semiotics of Law* 163–74; "Legal Deduction, Legal Predicates, and Expert Systems" (1992) V/14 *International Journal for the Semiotics of Law* 181–202 and "A Deductivist Rejoinder to a Semiotic Critique" (1992) V/14 *International Journal for the Semiotics of Law* 215–24.

occurred in a concrete case at a specified time in a way that materially involves the accused person or defendant. So the relevant normative consequence "NC" ought to be implemented as demanded. This is the standard legal syllogism[33] variously embodied in criminal or civil pleading and procedure,

But the conclusion is only as good as the premises, and these may be problematised. The challenge can be on proof, on characterisation, on interpretation, on relevancy (one, some, or all of them). But the idea of the rule of law that has been suggested here insists on the right of the defence to challenge and rebut the case against it. There is no security against arbitrary government unless such challenges are freely permitted, and subjected to adjudication by officers of state separate from and distanced from those officers who run prosecutions. In private law litigation, a similar requirement appears in the need for visible impartiality of the judge.

After hearing evidence and argument, the court must decide. In deciding matters raised in the problems of characterisation or of interpretation, or of relevancy, the court may find it necessary and proper to develop a new understanding of the law, set a new precedent, that may confirm or qualify prior understandings. At the end, the case is either dismissed as inconclusive, the defendant being absolved, or some order is made by the court and justified in the light of law as clarified through resolution of the problems posed. And then there is in effect a concluding syllogism. But it is rarely if ever identical with the starting syllogism. It is a new defeasible certainty that has emerged from posing problems about the old defeasible certainty and resolving them by rational argument.[34] From confronting law's arguable character, we move to restating a new putative certainty after admitting and dealing with doubts about the old.

In the upshot, it has to be recognised that the original representation of the "Rule of Law" as antithesis to the Arguable Character of Law was a mis-statement in the emphasis it gave to certainty in law. Whatever care is lavished on the source materials of law by legislators, drafters, or judges writing opinions that attempt to state a holding or *ratio* with exemplary clarity, the rule statements these yield as warrants for governmental action aimed at vindicating public or private right are always defeasible, and sometimes defeated under challenge by the defence. Law's certainty is then defeasible certainty. Its being so is not, after all, something that contrasts with the "Arguable Character of Law", but some-

[33] Compare Kennedy, *Adjudication*, n. 7 above, pp., 101–4; Burton, *Introduction*, n. 29 above, pp. 43–58.

[34] On defeasibility, see H. L. A. Hart, "The Ascription of Responsibility and Rights" (1948–9) 49 *Proceedings of the Aristotelian Society* 171–94; disowned by Hart in *Punishment and Responsibility* (Oxford: Clarendon Press, 1968), Preface; but see also G. P. Baker, "Defeasibility and Meaning", in P. M. S. Hacker and J. Raz (eds.), *Law, Morality and Society* (Oxford: Clarendon Press, 1977)m pp. 26–57. Compare N. MacCormick, "Law as Institutional Fact", Edinburgh University Inaugural Lecture No 52 (1983); (1974) 90 *Law Quarterly Review* 102–29; now in N. MacCormick and O. Weinberger, *An Institutional Theory of Law* (Dordrecht: D. Reidel Publishing Co, 1986), ch. 2, now superseded by MacCormick "Defeasibility in Law and Logic", in Z. Bankowski and I. White, *Informatics and the Foundations of Legal Reasoning* (Dordrecht: Kluwer Academic, 1995), pp. 99–117.

thing that shares an underlying ground with it. That ground is a conception of the rights of the defence built into the ideology of the rule of law in its guise as protector from arbitrary action by governments.

To conclude, then, we may hope that the rhetoric of the present essay, starting with an apparent opposition of ideas expressed in two competing commonplaces or argument bites, succeeds in its attempt to reconcile them by unravelling their real point in the legal context. There is a risk of misunderstanding the "Rule of Law" as an ideal taken in isolation. Then, perhaps, we stress its more static aspects, centring on legal certainty and security of legal expectations. But it has a dynamic aspect as well, centring on rights of the defence, and the importance of letting everything that is arguable be argued so long as a defender—or plaintiff—wishes to test out a reasonable legal argument. In this dynamic aspect, the argumentative character of law is no antithesis of the "Rule of Law", but one of its components.

9

Utopia and the Rule of Law

CHRISTINE SYPNOWICH[1]

Why should we follow legal procedures which might constrain our pursuit of justice? This question is at the heart of the rule of law. The idea of the rule of law is that the substance and application of law should meet certain formal standards. The purpose of these standards is to protect citizens who are subject to the law from arbitrary or unpredictable treatment. On this view, legality is not a mere means, whose purpose is only to facilitate efficiently whatever ends a legislature pursues; rather, the rule of law checks our political ventures, and can restrict the means we deploy to further our ends. As such, the rule of law is often viewed with suspicion, both by those who seek social change, and those who aim to prevent it.

In this essay I assess the suspicions of those who seek social change. These suspicions amount to an argument from utopia against the rule of law. The word "utopia" may seem to prejudice the case, since we usually think of utopia as by definition an unattainable ideal. But I think we should take seriously the idea that an ideal society involves a rejection of the rule of law, not in order to dismiss the idea of ideal societies, but in order to consider how the rule of law does or does not contribute to their pursuit.

The radical hostility to the rule of law involves a number of overlapping arguments which I will group into three categories. The first is an "equality critique" rooted in the socialist tradition. Socialists have traditionally been hostile to the general idea of law, taking it be an instrument for property relations which would cease to exist in a propertyless utopia, but they have been especially critical of the formalist obstacles posed by the rule of law to the pursuit of substantive justice. The idea that there could be a utopian society which transcends the impartiality and proceduralism of the rule of law has found adherents in new radical positions. This is the source of a second set of suspicions, comprising a "difference critique". The idea of a feminist ethic of care, for example, suggests that we draw on the different experiences of women, and import familial relations of intimacy into the public domain to replace impartial legal rules. Related to this are arguments on behalf of minority groups which call for disclosures of

[1] Thanks to David Dyzenhaus for his help and support throughout the writing of this essay, to David Bakhurst for valuable comments, and to the Social Sciences and Humanities Research Council of Canada for generously funding my research.

relations in the hitherto private and the public recognition of cultural differences. Finally, there is a third set of arguments which point to a fundamental tension between the rule of law and democracy, taking issue with the idea that the people's will should be constrained by procedural rules. This "democracy critique" is not unique to left-wing critics of the rule of law, but it is perhaps the most powerful when linked with a programme of reform and social justice. I will critically assess these positions to show the value of the rule of law for egalitarian politics, and moreover, its contribution to more general reflections on the pursuit of utopia.

I. THE RULE OF LAW

What is the rule of law? The term is a curious one, used to refer to a diversity of practices. First, in its most minimal form, it refers to rule by law rather than force. Agents of the state must act according to law. The contrast with force leads, however, to a second definition, which focuses on the obedience of subjects to the law, or the suppression of lawlessness. This "law and order" definition is often used to justify absolute obedience to the state and the limitless authority of the state to eliminate disobedience. It underlay the efforts of the British government under Margaret Thatcher, for example, to override established judicial procedures in order to combat Irish terrorism. In Canada, the federal government's efforts to persuade the Supreme Court that Quebec secession violates the rule of law had something of the same aspect of emphasising the compulsion of law. But the law and order view runs afoul of a third definition, which takes the rule of law to be a set of procedural constraints. This idea elaborates that minimal notion of law in contrast to force, and it also expresses the idea with which I introduced our topic, that law regulates the pursuit of our ends. On the procedural view, law must be so framed as to ensure that the individual is in a position to obey it.

At their most basic, the terms the rule of law, due process, procedural justice, legal formality, procedural rationality, justice as regularity, all refer to the idea that law should meet certain procedural requirements so that the individual is enabled to obey it. The instruments for effecting this idea are various: a written constitution and a constitution such as Britain's historic Magna Carta and Bill of Rights, and the plethora of principles established by the common law. These requirements are that first, law be general. That is, law must take the form of *rules* which are by definition directed to more than a particular situation or individual. The rule of law also requires that law be relatively certain, clearly expressed, open, and adequately publicised. A legal system should be internally consistent, so that particular laws do not conflict with each other. In addition, law must be prospective, directed only at behaviour which takes place after it is enacted. Retroactive law, which addresses actions taken before the law was made, prevents people from taking the requirements of the law into account

when planning their affairs. The practical effect of these principles is to set limits to the discretion of legislators, administrators, judges and the police. The rule of law dictates that these different aspects of governance are kept separate, so that political interference in legal affairs, for example, and the arbitrary power which is its result, are impermissible. The rule of law also aims to regulate internally the sphere of action of each of these functions of government. Detention without charge, arbitrary decrees, conviction without sufficient evidence, unduly harsh punishment: all would constitute violations of the standards of consistency and coherence integral to a system of law. Law and legal conduct which fail to meet these standards fail in their very function as constituents of a code of behaviour which individuals can consult when deciding how to act.

The protection of individual privacy is essential to the idea of rightly ordered legal institutions. Because the state by definition enjoys sovereignty over its citizens, and thus possesses a monopoly of putatively legitimate coercion, the state's invasions of privacy are potentially very harmful to those concerned. There are a number of ways in which the state's power is checked for the sake of individuals' privacy. Property rights, whilst usually conceived in terms of market exchanges and the accumulation of capital, also refer to the more mundane but highly prized personal property which the state cannot invade or appropriate except under very special circumstances. Rights to freedom of conscience, opinion, association and expression involve respect for the citizen's privacy from the state. Legal rights which protect the individual from arbitrary arrest, lack of legal counsel, or an unfair trial, also provide the means for demarcating the private realm from the public. These rights can be spelled out in constitutions or statutes or embedded in common law, or assured by some combination of the three.

The rule of law plays a vital role in the protection of privacy because it ensures that law's intrusions on the private should not be arbitrary or unpredictable.[2] We need to know where the boundary between public and private is drawn in order for our privacy is to be protected; it is thus essential that the actions of public officials are regulated so that the public-private distinction is a reliable one.

It would be an overstatement to say that there would be no privacy without the rule of law; after all, just as a practical matter, seclusion can never be entirely obliterated, even in putatively "totalitarian" societies. Moreover, people probably retreat into the private all the more in a society characterised by procedural injustice. Consider the symbolic importance of gathering in the kitchen, "*na kukhne*", for political conversation in Soviet Russia, even though Russians were acutely aware that surveillance could intrude the home (it was a common belief, no doubt erroneous, that all telephones were bugged).[3] Nonetheless, insofar as

[2] This discussion of privacy draws on material from "The Civility of Law: Between Public and Private", in Maurizio Passerin d'Entreves and Ursula Vogel (eds.), *Public and Private: Legal, Political and Philosophical Perspectives* (London: Routledge, forthcoming).

[3] I am grateful to David Bakhurst for this example.

privacy is particularly significant as a zone of non-interference from the state, it is incontrovertible that the rule of law is essential for such privacy.

For John Locke, one of the rule of law's first exponents, the chief advantage of civil society over the state of nature is the assurance of "established, settled, known law", applied by a judge who is both "known and indifferent", who does not produce judgments that are "varied in particular cases, but to have one rule for rich and poor, for the favourite at court, and the country man at plough".[4] The impersonal tenor of Locke's ideal persists in what is probably the most pervasive image of justice, that of a blindfolded woman weighing scales, as though it is the scales, rather than the woman, who renders judicial decisions.[5] In fact, of course, the law is drafted, and subsequently applied, by people. It is as people, with particular interests, needs and aims, that we are motivated by the rule of law and its dictates. We look to the rule of law to rein in, or check, the foibles of the human in the administration of justice. The rule of law seeks to render legality impartial, abstract, neutral, general, universal: all of these words indicate something of the transcendent, even if it is a procedural rather than substantive set of standards which would transcend the hurly-burly of human affairs.

II. THE EQUALITY CRITIQUE

Socialists have argued that the rule of law is at odds with the pursuit of substantive justice.[6] Interestingly, this argument underlies, not just the left-wing critique of the rule of law, but right-wing critiques of the welfare state. F.A. Hayek argues that the rule of law's focus on procedural rather than substantive justice keeps state interference to a minimum, so as to enable individuals to make private economic decisions which are the mainstay of capitalist efficiency. The rule of law requires that legislation do no more than provide a formal framework for private initiatives, and it is most vulnerable when the welfare or socialist state threatens to "engulf the private sphere". For Hayek, the redistribution of wealth requires discretionary powers on the part of government which makes for arbitrariness in the law.[7] The classical socialist antipathy to the rule of law thus

[4] Locke, *Second Treatise of Government*, (C.B. Macpherson (ed.), Indianapolis: Hackett, 1980), ch. XI, p. 75.

[5] Indeed, that justice is represented by a woman in this early symbol, before women can vote or own property, let alone act as lawyers or judges, attests to the irrelevance of the corporeal person who holds the scales. Judith Shklar makes much of an earlier image of justice in Giotto's picture in the Arena Chapel in Padua. *La Giustizia* is not blindfolded, but her benign, expressionless face does suggest, Shklar says, that "she may not be a real person at all": *The Faces of Injustice* (New Haven: Yale University Press, 1990), p. 103.

[6] This section draws on material from chapter 3 of my book, *The Concept of Socialist Law* (Oxford: Clarendon, 1990).

[7] *Road to Serfdom* (London: Routledge and Kegan Paul, 1971), pp. 56–9; *Constitution of Liberty* (London: Routledge and Kegan Paul, 1960), pp. 214–16.

confirms Hayek's view that efforts at equality run afoul of the rule of law's emphasis on procedural constraints.

Judith Shklar has expressed some impatience with the role of the rule of law as "a football in a game between friends and enemies of free-market liberalism".[8] But that Hayek and his socialist critics concur that the rule of law cannot be married with egalitarian aims suggests that such a view has considerable force, particularly troubling if we are concerned both to promote equality and to follow the dictates of the rule of law.

The socialist's "equality critique" can be broken down into three complaints. The first points to the rule of law's connection with privacy and, by implication, private property. The rule of law's focus on certainty and predictability is at odds with redistributive aims, which inevitably disrupt the status quo and thereby the certainty of the propertied; procedural justice thus ensures that unjust economic power goes unquestioned under the guise of guarding the domain of the private. Indeed, some go so far as to say that the rule of law's historical role in safeguarding property relations undermines its claim to impartiality.

Certainly if intrinsic to the idea of the rule of law is that its concern for privacy dictates that private property be immune to interference, then the conflict with arrangements that seek to redistribute wealth is obvious. Historically the idea of personal freedom has been couched in terms of implying a right to the accumulation of private property. Whilst some idea of a personal domain is essential for privacy to be respected, such a domain need not be owned. A peeping tom invades the privacy of a guest or housesitter as much as that of the owner or tenant. As Thomas Scanlon observes, "ownership is relevant in determining the boundaries of our zone of privacy, but its relevance is determined by norms whose basis lies in our interest in privacy, not in the notion of ownership".[9]

Although ownership per se is not essential to privacy, titles to property are a way of expressing the claim to privacy. Private property is an important means of constituting the inviolable zone or territory that attaches to the person whose privacy is at stake. We might elect to protect privacy by means of private property, not just in the form of personal effects, but also, perhaps, a place of residence. Privacy is compatible, however, with serious restrictions on the extent of ownership. Large property claims, that is, claims to capital, returns on investment, income, and so forth, are not necessary to secure privacy, nor are they even rightly characterised as private. Indeed, what distinguishes these larger forms of property is their public nature, in the way they are produced and exchanged, and in the interests they affect. The historical links between privacy and the inequalities of private property can be severed because there is no contradiction between respecting individuals' privacy whilst redistributing wealth

[8] J. Shklar, "Political Theory and the Rule of Law", in A. Hutchinson and P. Monahan (eds.), *The Rule of Law: Ideal or Ideology* (Toronto: Carswell, 1987), p. 16.

[9] Thomas Scanlon, "Thomson on Privacy" (1975) 2/2 *Philosophy and Public Affairs* 318.

to further equality (though there may be practical matters about redistribution that involve tampering with the private). Wealth might even be redistributed to further equality of privacy.[10] Given the importance of privacy for individuals' well-being, egalitarians should avoid latching on to right-wing views about the redistribution of wealth eroding the protection of privacy.

This conclusion is further supported by a body of literature attacking Hayek's argument about the rule of law's incompatibility with the welfare state. These rebuttals take a variety of forms. One tack is to concede the conflict between the rule of law and redistribution but to insist on the inevitability of compromise: the rule of law is only one value among many in a legal system. Another is to repudiate the existence of a special conflict in this case, since redistributive policy can be stated in a clear and principled way just like other policies, and its application in concrete situations is the fate of all general rules. In any case, it seems inconsistent to insist on the procedural justice of the rule of law on the one hand, and on the other, to hold that the rule of law entails a particular conception of the substance of law.[11]

A stronger argument for the compatibility of the rule of law and the redistribution of wealth can be made which suggests that the rule of law would be enhanced by social and political relations of equality. Where property is more equally distributed, there would be less likelihood of social panics about crimes against property, and the unpredictablity they bring in their train. Access to good legal counsel would be more equally distributed, in keeping with the principle of generality where the law is no less applicable to, as Locke says, the "favourite at court" than the "countryman at plough". And the composition of the judiciary, both demographically and ideologically, might be less skewed in favour of the wealthier social classes, also contributing to greater certainty and generality in the law.

The equality critique's second objection to the rule of law is the ideological role played by procedural justice. On this view, not only is procedural justice protected at the expense of substantive justice, but the former provides an ideological justification for the absence of the latter. The rule of law, trumpeting the morality of procedures, thereby occludes the issue of a more contentful morality of equality. Connected with this is a third objection, that the rule of law fosters a legalism about justice so that rule-following takes the place of a concern with outcomes and substance, and moreover, takes the place of concern more generally, in the sense of fellow-feeling and community.

[10] In a path-breaking argument, Charles Reich maintained that welfare entitlements should be seen as a species of property rights, bolstering the idea that an equitable distribution of personal resources facilitates equal enjoyment of privacy: see Reich, "The New Property" (1964) *Yale Law Journal* and C.B. Macpherson, "Human Rights as Property Rights", in *Rise and Fall of Economic Justice and Other Essays* (Oxford: Oxford University Press, 1985).

[11] See Joseph Raz, *The Authority of Law* (Oxford: Clarendon, 1979), p. 228; Julius Stone, *The Province and Function of Law* (Sydney: Maitland Publications, 1950), pp. 262–3; D.J. Galligan, *Discretionary Powers* (Oxford: Clarendon, 1986), p. 205; Harry Jones, "The Rule of Law and the Welfare State" (1958) 58/143 *Columbia Law Review* 150.

The ideological aspect offers, in fact, a promising strategy for confirming my argument that the rule of law has value for socialism. As E.P. Thompson has famously argued, if the rule of law is to have an ideological function, camouflaging substantive injustice, it must further values which are first, in fact valuable, and secondly, capable of being realised, in however partial a form. If procedural justice was a complete sham, no ideological purpose would be served.[12] An analogy can be made with good manners and cruelty. If someone's cruelty is disguised by their good manners, then first, the good manners must have some genuine existence, and secondly, they must have value. If we could eliminate the cruelty, therefore, we might still want the good manners. Analogously, if we could eliminate substantive injustice, we might still want to hold on to procedural justice. The role of the rule of law in protecting privacy, as we have seen, is not dependent on a regime of private property; the value of privacy transcends its historic, inegalitarian context. Indeed, the rule of law's check on arbitrariness, I suggested, might be particularly well deployed in the private domain where arbitrary social power, economic in origin, needs to be checked and regulated.

The charge of legalism, too, looks easier to meet, if it turns out that adherence to procedures is not necessarily at odds with concerns for substantive justice. Nonetheless, the worry about legalism may persist, however much it turns out that procedural rules are consistent with substantive justice. This is because at issue is the utopian vision that accompanies ideas of substantive justice. This vision involves a concern that the proceduralist ethic dictates we sacrifice communitarian values for the sake of fidelity to rules. Proponents of the equality critique have in mind relations of community and fellow-feeling which would characterise an ideal socialist society. But it is not just socialists who adduce such values. These values are central to the difference critique.

III. THE DIFFERENCE CRITIQUE

It might be said that in our culture, the most obvious instance of a group different from the "norm" is women. The significance of the differences between the sexes, however, is a matter of controversy for feminists and sexists alike. Feminists have traditionally pointed to liberal values such as procedural justice in order to argue that women be treated as the equals of men, and their differences from men be either eliminated or downplayed. Recently there has been considerable criticism from feminist quarters about this strategy. According to Carol Gilligan, women have a psychological history different from men's which produces a set of values in opposition to a proceduralist ethics:

[12] E.P. Thompson, *Whigs and Hunters: The Origins of the Black Act* (New York: Pantheon, 1975), pp. 264–5.

> "Since the reality of connection is experienced by women as given rather than freely contracted, they arrive at an understanding of life that reflects the limits of autonomy and control . . . While an ethic of justice proceeds from the premise of equality, that everyone should be treated the same, an ethic of care rests on the premise of nonviolence, that no one should be hurt."[13]

An ethic of care attends to the particular, perceives the connectedness of human beings with each other, takes a posture of care and nurture, and is concerned about outcomes: at odds with the rule-bound, formalistic procedures of impartial justice. The idea of a distinctive female voice of care is akin to the idea of "maternal thinking" developed by Sara Ruddick, which she offers as a feature of women's "cultures, traditions, and inquiries which we should insist upon bringing to the public world".[14] The ethic of care, it is argued, will better meet people's needs, and it will also foster a community of fellow-feeling, sympathy and mutual regard which the rule of law, in its impersonal, abstract approach, cannot supply. The ethic of care thus invokes some of the socialist ideas about a society of fellowship and community, beyond justice. This ideal is connected to the hostility to legalism, since it involves a notion of sympathetic and sociable persons who relate to each other transparently, unmediated by political and legal institutions, unbounded by impersonal rules. Robin West argues that modern jurisprudence is masculine in large part because of its commitment to the rule of law, a concept which is counter to women's "material and existential circumstance" of connection and intimacy.[15] Indeed, some feminist legal theorists have sought to revise concepts of law in light of this anti-legalism. Jennifer Nedelsky, for example, has argued that we "reconceive rights as relationship" to better reflect "the ways in which our essential humanity is neither possible nor comprehensible without the network of relationships of which it is part".[16]

The case is further supported by arguments made on behalf of "difference" more generally. These arguments point to the disadvantaged position of members of minority groups of a variety of kinds, ranging from race, sexual orientation, ethnicity, or disability, to repudiate both privacy and proceduralism. Iris Marion Young, the most prominent exponent of this view, targets the "civil public" which "expresses the universal and impartial point of view of reason, standing opposed to and expelling desire, sentiment, and the particularity of needs and interests".[17]

The difference critique, in its rejection of the ideal of impartiality intrinsic to the rule of law, has had considerable influence in recent debates in political

[13] C. Gilligan, *In A Different Voice* (Cambridge, Mass.: Harvard University Press, 1982), pp. 172–4.

[14] S. Ruddick, "Maternal Thinking", (1980) 6/2 *Feminist Studies* 345.

[15] R. West, "Jurisprudence and Gender" reprinted in Patricia Smith (ed.), *Feminist Jurisprudence* (Oxford: Oxford University Press, 1993), pp. 520–2.

[16] J. Nedelsky, "Reconcieving Rights as Relationship", (1993) 1/1 *Review of Constitutional Studies*, p. 12.

[17] I. Young, *Justice and the Politics of Difference* (Princeton: Princeton University Press, 1990), p. 108.

theory. It rests, however, on some mistaken assumptions about the rule of law, and a naivete about the kinds of social relations that would obtain in its absence.

First, let us consider the role of the rule of law. It is important that we distinguish between the real, historical link between the rule of law and relations of inequality, and a supposed, conceptual link. Like the socialist argument for equality, the difference argument risks taking one for the other. The incontrovertible record of disadvantage suffered by women or members of cultural minorities in modern liberal societies is, I think, a function of a failure to live up to the demands of proceduralism, not an indication of failure in the concept. Indeed, it is to the concept that we turn in order to perceive, argue against, and remedy cases where women and members of minority cultures have not been treated as equals before the law.

The second challenge of the difference critique lies in its proposals for alternatives such as an ethic of care, or a politics of recognition. At issue is the idea that we shake off proceduralism in favour of more transparent social relations. Again there is some overlap with the socialist ideal of a society, rid of private property and social classes, where citizens relate to each other on the basis of direct expressions of need, fellow-feeling and community. The difference critique is launched, however, on behalf of a constituency markedly unlike the have-nots of the socialist critique. Here economic disadvantage may be an effect of the oppression at issue, but it is cultural factors which are the more fundamental source. This means that, paradoxically, it is the difference constituency in particular which stands to lose in the face of the repudiation of the rule of law. For it is these groups whose disadvantage resides in a culture of condescension, intolerance or even hatred, to which an ethic of proceduralism is an important antidote. And there is a cultural lag on issues such as ethnicity, sex or race, so that despite legal advances, members of minority groups still face distressing behaviour or situations. Members of these groups may be further disadvantaged by the removal of procedural justice and the embrace of a more intimate, revelatory kind of justice. Doing away with procedures presupposes citizens' care and concern to be adequate to the task of attending to members of groups beset by stigma and prejudice; but it is precisely the existence of such stigma and prejudice which indicates the dubious efficacy of care and concern. As Patricia Williams notes, the disadvantaged person seeks, not intimacy, but the status of the "bargainer of separate worth, distinct power, sufficient rights".[18] However contrary to their intentions, there is in the difference proposals a disregard for the risks of disclosure suggestive of the controversial cases of non-consensual "outing" of gays.

Erving Goffman's work underscores the risk we all feel, however subtly, in the public domain, our vigilance in warding off danger, however minor, in our encounters with others. In a sense, it is intrinsic to society that the public is not

[18] Patricia Williams, *The Alchemy of Race and Rights* (Cambridge, Mass.: Harvard University Press, 1991), pp. 146–8.

characterised by loving care. This is not because the public is a Hobbesian world devoid of consideration for others, but because it is consideration of a distanced kind, based on the "intricacies of mutual trust presupposed in public order". In public we seek to preserve the integrity of what Goffman calls a personal "surround"; something which is threatened not just by the pickpocket or flasher, but by the overly solicitous or familiar.[19] Care is inappropriate in this domain because it lacks the background conditions of intimacy and knowledge of the other. We lack, that is, the depth of feeling for strangers that care requires, and we are too unacquainted with each other to know what care would consist of. Moreover, in a society of inequality and discrimination, care may be an unlikely prospect. In Ibsen's famous play, lacking "a room of her own", Dora flees the "Doll's House" precisely to be rid of the suffocating, ill-judged care of her patronising husband. Care is of no help here; indeed, it is too implicated in conditions of oppression. Even were we to diminish greatly inegalitarian attitudes, there would still be conflicts of interest that render care difficult to muster in matters of public concern. Moreover, if we take seriously the Marxist insight about the structural conflicts of interest between people with different material positions, then care in the abstract looks idealistic indeed. In such a context, to contend, as does Nedelsky, that legal institutions such as rights somehow be the expression of care and relationship, risks a naive understanding of the nature of the public domain.

IV. THE DEMOCRACY CRITIQUE

We have answered the socialist and difference critiques with some arguments for the role of the rule of law in remedying disadvantage. But there will linger some populist misgivings about a focus on procedures and rules. Here radical objections to the rule of law, exotic though they may seem, have a life in mainstream debates. This is the populist worry about legalism, which arises in confronting the connection between the substance and form of law, and in particular between the substance of law as it is enacted by democratically elected governments and procedures enforced by unelected judges. Does the rule of law give priority to procedures willy-nilly, with no regard for the will of the people? In short, is democratic government compatible with the rule of law?

In A.V. Dicey's famous doctrine, the supremacy of Parliament is set out as the companion principle to the rule of law.[20] That the judiciary should defer to the commands of an elected Parliament can be portrayed as essential to the idea that the individual has control over his or her affairs, knowledgeable of what the law requires, and capable of seeking to change the law. There is, however, an obvious tension between the two principles, since if the rule of law is to check

[19] Erving Goffman, *Relations in Public* (New York: Harper and Row, 1971), pp. 331, 252.

[20] A.V. Dicey, *Introduction to the Study of the Law of the Constitution* (London: Macmillan, 1964), p. 406.

legislation in light of procedural standards, the judiciary must be capable of challenging Parliament. As T.R.S. Allan puts it, "the traditional role of the common law in defence of justice and liberty , as those ideals have been understood, is radically inconsistent with a notion of unlimited legislative supremacy".[21]

The difficulty of reconciling procedural restraints with parliamentary supremacy has given rise to a polemical attack on the rule of law by Allan Hutchinson and Patrick Monahan, who claim that:

> "the Rule of Law is a politics with limited scope for popular participation and control. It cramps and compresses the ability of individuals to debate and define the conditions of their communal life. In attempting to avoid the tyranny of the majority, it mistakenly embraces a doctrine of expertise and dependency which carries with it a subtle, yet despotic dominion of its own."[22]

For Hutchinson and Monahan, a commitment to democracy means refusing the role of the judiciary as a check on the community's policies. The "greatest possible engagement by people in the greatest possible range of communal tasks and public action" is incompatible with deferring to legal experts. Hutchinson and Monahan connect their democratic argument against the rule of law to the equality argument for intervening in the market and the difference argument for an alternative ethic of care and community. Democracy is the means by which a community refuses the model of the individual as an isolated, self-interested holder of property and with which it expresses a culture of mutual recognition. They thus contend that any politics which seeks to develop democratic community and social change has no use for the rule of law. Insofar as the community's democratic decision-making needs to be regulated, this can emanate from the principles of democracy itself, which guarantees and extends free elections, debate and assembly.[23]

Hutchinson and Monahan's position relies on a crude depiction of the rule of law as imprisoned within the historical circumstances of its origin, a view which we have grounds to reject in light of our discussion of the equality critique. Like the socialist antipathy to proceduralism, their view relies, paradoxically, on a Hayekian framework. This is because they actually subscribe to the view of democracy Hayek impugned, as being instantiated in the majority rule of a community to which individuals must submit. But there have been subtle arguments to suggest judicial interpretation in fact contributes to democratic government. Thus, for example, Jeffrey Jowell notes that the rule of law is concerned to protect both individuals and groups from being deprived of their rights without the opportunity to defend themselves. The idea that one ought to have the possibility of expressing one's views, to participate in decisions about the exercise of societal power, diminishes the contrast between the rule of law and democracy.

[21] T.R.S. Allan, *Law, Liberty and Justice* (Oxford: Clarendon, 1993), p. 17.

[22] A. Hutchinson and P. Monahan, "Democracy and the Rule of Law" in Hutchinson and Monahan (eds.), *The Rule of Law: Ideal or Ideology*, n. 8 above, p. 111.

[23] Ibid., pp. 119–22.

Insofar as the rule of law disables government from abusing its power, it represents the citizenry's interest in good representation.[24] As Ernest Weinrib argues, essential to the rule of law is the requirement that law be intelligible: "the form provides the deep structure which is realised in the legal content and through which the content can be understood".[25] Proceduralism thus also serves democracy insofar as it requires that the people understand how they are being governed. Moreover, the rule of law's role in disabling government from abusing its power serves to represent the citizenry's interest in good representation.[26] As Stephen Sedley retorts, "ministers are no more elected than judges are"; an important role for the courts is thus to safeguard the rule of law in the face of the otherwise unaccountable abuses of ministerial power.[27]

These are, I think, good reasons for conceiving of the rule of law as contributing to democratic government. However, they should not be taken to mean that the ideals of the rule of law are synonymous with democratic ideals. Something like this assumption underlay some arguments launched in response to Margaret Thatcher's measures to centralise government power and restrict civil liberties. In their grim 1990 study of civil liberties in Britain, for example, Ewing and Gearty conclude with a call for a democratic culture, electoral reform and checks on executive power.[28] It was an understandable response to a Tory government's disregard for procedural justice to call for different structures of governance, if not a different governing party. But a different set of representatives or a broader mandate do not necessarily make for greater respect for liberty under law.[29] The Labour government under Tony Blair has, for example, stepped up restrictions on individual liberties in response to Irish terrorism.

The complex question of democracy's relation to the rule of law is well-illustrated by David Dyzenhaus's study of the recent legal hearings of the Truth and Reconciliation Commission in South Africa. He notes that in the apartheid era, supremacy of Parliament was a favourite rationale amongst South African judges for applying inequitable legislation. They could abrogate responsibility for assessing an unjust law or the unjust application of law by claiming that they had to defer to the will of the legislature from whence law had come. Indeed, some went so far as to interpret legislation in light of intentions

[24] J. Jowell, "The Rule of Law Today", in J. Jowell (ed.), *The Changing Constitution* (Oxford: Clarendon, 1994), pp. 71–2.

[25] Ernest Weinrib, "The Intelligibility of the Rule of Law", in Hutchinson and Monahan (eds.), *The Rule of Law: Ideal or Ideology?*, n. 8 above, p. 81.

[26] Jowell, "The Rule of Law Today", n. 24 above, at 72.

[27] S. Sedley, "The Common Law and the Constitution", *London Review of Books* (May, 1997), p. 11.

[28] K.D. Ewing and C.A. Gearty, *Freedom Under Thatcher: Civil Liberties in Modern Britain* (Oxford: Clarendon, 1990), pp. 255–6, 275.

[29] For a sanguine assessment of the relative merits of the different political parties on issues of proceduralism and civil liberties in Britain, see Cosmo Graham and Tony Prosser, *Waiving the Rules: The Constitution Under Thatcherism* (Milton Keynes: Open University Press, 1988), pp. 174–93.

deduced from other statutes or government debates. According to Dyzenhaus, the supremacy argument fails because the Parliament in question was insufficiently democratic: in apartheid South Africa, suffrage was based on race, and hence Parliament was unworthy of supremacy.[30] The implication of this view is that Parliament's supremacy provides a ground for judges to prescind from deciding on the procedural morality of law, so long as Parliament is rightly constituted.

It is of course galling indeed for judges to invoke the idea of deferring to the will of the people in an undemocratic society such as apartheid South Africa. It is not clear, however, that a philosophy of law based on the modest principles of the rule of law contains the resources for requiring of Parliament that it be democratic. Dicey, writing in the nineteenth century, certainly did not make universal suffrage a condition of his principle of Parliamentary supremacy. Perhaps we could say that law rightly composed is now understood as law that is democratically made; that law must emanate from a fully representative legislature is now one of the procedures intrinsic to the rule of law. Controversy over the adequacy of the representativeness of the legislature may remain, however. Consider the protests of the Thatcher era that a prime minister could obtain power with a minority of the votes cast, that the "first past the post" electoral system made for illegitimate mandates, that the executive's unchecked power was tantamount to authoritarian: all raising doubts about the claim to democracy in what is taken as the model of parliamentary systems.[31]

In any case, even if we concede the idea that Parliament rightly composed is supreme, we run the risk of forfeiting important principles of the rule of law. Indeed, these principles are at the heart of Dyzenhaus's argument in favour of the idea of the common law as the repository of procedural justice. It is on these grounds he finds South African judges guilty of a dereliction of duty where rules about inadmissable evidence or the right to a fair trial were disregarded.[32]

These juridical principles are particularly pressing in the new, representative South Africa. The arrival of democracy in South Africa is a wonderful thing, but it does not mean that the rule of law should play a diminished role as a procedural watchdog. Indeed, we might worry that, in the process of transforming South Africa, the idea that the means of doing so might be regulated by the rule of law is particularly vulnerable. In a footnote early in the book, Dyzenhaus notes that the current government has been criticised by human rights activists for their initiatives on maximum security prisons and the granting of bail.[33] South Africans, the majority of whom were disenfranchised and brutalised for so long, may well be impatient with procedural niceties in the face of the imperative for social change, or a desperate need for stability and order. Certainly this

[30] David Dyzenhaus, *Judging the Judges, Judging Ourselves: Truth, Reconciliation and the Apartheid Legal Order* (Oxford: Hart Publishing, 1998), p. 76.

[31] Ewing and Gearty, *Freedom Under Thatcher*, n. 28 above, pp. 255–7.

[32] Dyzenhaus, *Judging the Judges*, n. 30 above, pp. 50, 71.

[33] Ibid., p. 8.

has been the case for past radicals, be they on behalf of the dispossessed or the possessors, Bolsheviks or Thatcherites. There is thus some merit in the view, rejected by Dyzenhaus, that the rule of law be conceived of as an "anti-politics" in some sense.[34]

V. MORALITY AND POSITIVISM

These considerations about democracy and proceduralism point to the difficult question of the place of the rule of law in a society dedicated to the pursuit of justice and the dignity of the individual, the society that we hope is emerging in today's South Africa. That is, whilst the rule of law is obviously absent or undermined in dystopia, what role might it have in a society with utopian aspirations? We might assume that these societies are by definition intent on fulfilling the rule of law. Or, as I worried earlier, we might presume that they need not be too mindful of it. But the rule of law requires that the pursuit of our ends, no matter how sound and just those ends might be, meet certain moral standards. This means we must be prepared to see our ends thwarted if they are pursued in a procedurally deficient way. And in so doing, we might come to see that our ends were compromised by the procedural deficiences of our means. Historical experience confirms this. The roots of Stalinist jurisprudence, after all, were in a revolution which was animated by powerful ideals of equality and freedom. And we are critical of that revolution in large part because it excluded the proceduralism of law, both as a means and as an end.

How substantial is the morality offered by the rule of law? The philosophy of law has been riddled with a longstanding debate about whether morality is essential to the validity of law.[35] Positivists argue that law's sources reside in "social facts": rules internal to a system of law which specify what counts as law in that system. The sources of law are institutional, and have no necessary connection to moral rightness, however it may be defined. Anti-positivists take the view that law must meet some minimal moral criterion in order for it to count as law. The anti-positivist view takes various forms, from John Finnis's Catholic natural law, to liberal conceptions of constitutionalism such as that of Ronald Dworkin.[36]

Some positivists, such as Hobbes, went so far as to take the rule of law to mean, not standards that the law must meet, but the standard imposed by law that, regardless of its moral content, individuals must obey. Such authoritarian, "law and order" versions of positivism have been displaced in more recent

[34] Ibid., p. 22.

[35] Here I draw on my argument about the sources of law in chapter 2 of my book, *The Concept of Socialist Law*, n. 6 above.

[36] See J. Finnis, *Natural Law and Natural Rights* (Oxford: Clarendon, 1980), and R. Dworkin, *Taking Rights Seriously* (Cambridge, Mass.: Harvard University Press, 1978), pp. 160–1 and *Law's Empire* (London: Fontana, 1986), pp. 225, 400.

positivist theories (although they continue to have currency in practice, as many of our examples have shown). H.L.A. Hart conceded that morality might come to play where the law is uncertain or silent; the judge must draw on moral materials in the absence of reliable sources.[37] Joseph Raz offered some solace to the conscientious objector, holding that positivism distinguishes legal from moral obligations, freeing the individual to clear-sightedly opt for the latter where the law is immoral.[38] Nonetheless, the positivist idea that law is empty of any necessary moral content remains.

The positivist account of the sources of law seems inadequate, for two reasons. First, it fails to consider that the tyrannical system may not be law precisely because its procedures lack the moral dimension intrinsic to the procedural sources of valid law. And secondly, it fails to persuade us that the presence of moral evaluation in the assessment of the difficult or unclear case has no relevance in the certain one. Cases differ to the extent that the moral criteria in question is contentious, complex or difficult, but not to the extent that morality is absent or present. On the other hand, the thick conceptions of morality deployed by anti-positivists, such as some idea of natural law, set moral criteria which is intrinsically contentious. The problem is not just that natural lawyers do not all ask this question in the same way. We would expect natural law to be difficult to apprehend and its proponents to be divided about its content. What is troubling, though, is the very project of setting a moral standard that transcends the different moral situations which arise in a society, not to mention those posed by different cultures and historical epochs. Finnis is adamant that the common good is a standard of validity which cannot be "relative to the opinions and practices of a given community".[39] Dworkin's idea of "right answers" have a transcendent aspect too, insofar as they are conceived as emerging from a "single author" or "the community personified".[40]

The rule of law offers a resolution to this impasse, since it offers both an emphasis on the institutional basis of law, and a conception of the norms which these institutions protect. The term "procedural morality", which is often taken as synonymous with the rule of law, illustrates this idea that there is moral content in the requirement that law be framed in a way that renders it capable of being obeyed. In its very practice-based moral standards, the rule of law promises a minimum of justice, and inhibits, if not prevents, the use of legal institutions to promote injustice. The rule of law draws on institutions which are similar to that of the positivist idea of "social facts", without conceiving of the institutions as divorced from any ethical criterion. At the same time, the rule of law embodies a concept of justice like the natural lawyer's

[37] H.L.A. Hart, *The Concept of Law* (Oxford: Clarendon Press, 1972), pp. 119–32.
[38] Raz, *The Authority of Law*, n. 11 above, p. 42.
[39] *Natural Law and Natural Rights*, n. 36 above, p. 21.
[40] *Law's Empire*, n. 36 above, pp. 225, 400.

idea of law's moral authority, without drawing on an implausibly rigid or substantive conception of moral law. It is important, then, that we maintain a sense of the rule of law as a set of moral standards, without assuming these standards are met by the purposes of law, or reducing them to the principle that a democratic legislature should be law's source. It is for these reasons that I prefer to see the idea of the rule of law, not as the antidote to positivism,[41] but as a resolution to the impasse between natural lawyers and legal positivists.

The rule of law so understood moves us away from the contentious idea of neutrality, devoid of human concerns or norms; rather, the rule of law refers to several important principles. The most obvious is respect for individual liberty, which requires a stable and predictable context. Connected to this is respect for privacy, that the law cannot affect the individual's private domain except under certain regulated conditions. Equality is intrinsic to the rule of law insofar as it requires that all individuals are equally subject to the law. Democracy's relation to the rule of law is complex. I have argued that inherent in the rule of law is the idea that law must be accountable to people's interests, not in its mandate or source in a supreme parliament, but in its form. The judiciary interprets even the most democratically made law in light of principles which respect the equality of citizens before the law, and the liberty of citizens under the law.

The proceduralism inherent in the rule of law nonetheless means that these moral principles are principles that dictate restraint. Thus Hayek's message of limited government is unavoidable in some sense. Some have defended the rule of law in light of the value of its procedural virtues as possessing "substantive content".[42] It is true that seemingly empty formalistic rules are an important source of fairness, in their scrupulous attention to same treatment for all citizens. They are rooted in an idea of equal citizenship, an idea that has historically been widened and deepened to involve ideas of social justice and material equality. But to equate the rule of law with the substantive is I think misleading. The justification of the rule of law is founded in a political morality, but it is nonetheless a political morality of a narrow, circumscribed kind.[43]

This does not mean that we should find the rule of law in tension with social

[41] This is Dyzenhaus's approach in *Judging the Judges*, n. 30 above, and in his earlier work, *Hard Cases in Wicked Legal Systems: South African Law in the Perspective of Legal Philosophy* (Oxford: Clarendon, 1991).

[42] Jowell, "The Rule of Law Today", n. 24 above, at 74; Allan, *Law, Liberty and Justice*, n. 21 above, p. 46. But see Raz, *The Authority of Law*, n. 11 above, p. 211.

[43] This suggests that the Truth and Reconciliation Commission itself, which took on the task of "judging the judges", might be vulnerable in light of procedural considerations. If it is essential to the rule of law that the judiciary enjoy independence from other governmental bodies, then holding judges accountable to the Commission presents some problems, and Dyzenhaus's defence that the behaviour being held to account itself involved compromising the independence of the judiciary, does not vitiate the principle involved: see Dyzenhaus, *Judging the Judges*, n. 30 above, pp. 142, 172.

welfare or government intervention in the economy. As our examples of Thatcherism and apartheid justice make clear, the temptation to put aside the rule of law is not specific to the Left. And whilst the profusion of administrative tribunals since the emergence of the welfare state gives greater scope for discretion, discretion was always intrinsic to the *raison d'etre* of the rule of law. Were it not for the fact that the slogan "government of laws and not men" is impossible, that the law is in fact administered by human beings, there would be no need for the difficult task of regulating discretion.

It is thus unfortunate that even left-wing defences of the rule of law tend to reify formal law, accepting the idea that welfarist social policy is a legal oddity, requiring particularistic, outcome-oriented adjudication not in keeping with the generality required by the rule of law. William Scheuermann's vocabulary of "the norm" versus "the exception", for example, concedes too much to right-wing critics.[44] A general rule can specify the class accorded general treatment; otherwise, not just welfare law, but all law is at risk. Moreover, legality, in applying general rules to particular situations, finds its purpose in the "exceptions", the unanticipated. We invoke the norm to ensure legality is as measured and predictable as possible. We should not, however, in our haste to avoid radical utopianism fall prey to a conservative utopianism about the actuality of blindfolded justice under capitalism.

To conclude. I have argued that the rule of law is not at odds with concerns for substantive justice, but nor is it identical with such concerns. Its morality is not the morality of social justice, care or democracy. It is best thought, perhaps, as deploying an ethic of civility. Civility is the form of care appropriate for public life, distinct from the care of the intimate or private realm. Indeed, the care to which civility refers consists of the more muted sense of regard. The rule of law helps ensure that we are accorded worth and dignity in the domain of the public, that we are included and counted as citizens. But the rule of law also seeks to leave us unimpeded and unseen in our particular personal domains, according us respect as private persons. It directs government to be civil to its citizens, to treat citizens with a concern which is generalised and abstract, unintrusive and aloof. In its instantiation in the rule of law, civility retains that sense of safeguarding boundaries, reminding us that arguments for equality should ward off utopian appeals to transparency and community which dismiss the concern for regulating the community's interventions in the affairs of the individual.

Thus in perceiving the absence of the rule of law in dystopia, we should not assume its fulfilment is unproblematic in utopia. It is essential to the rule of law that it delineates our projects, checks and regulates them. If utopia is a world without rules, then so much the worse for utopia. Or rather, we should under-

[44] W. Scheuermann, *Between the Norm and the Exception: the Frankfurt School and the Rule of Law* (Cambridge, Mass.: MIT Press, 1995).

stand utopia as an aspiration central to politics, which is self-conscious as to its dangers and proper scope, and which seeks its own limitation in a legal and political order guided by the rule of law.

10

The Rule of Law Revisited: Democracy and Courts

ALLAN C. HUTCHINSON

Talk of the rule of law is never far from the centre of contemporary jurisprudential debate. Although the immediate focus of discussion may be more particular and focused, most juristic offerings are underpinned by a reasonably full set of operating assumptions about the scope and meaning of the rule of law. At the heart of the rule of law is the powerful idea that it is law that should govern society and not the arbitrary will of particular persons—*a government of laws, not persons*. As such, the rule of law seems to demand that a state puts in place and adheres to a body of definite rules amenable to an impartial application which all citizens have a moral obligation to obey and which can only be changed in accordance with the prescribed rules. If the rule of law is taken to work as a hedge against tyranny, then it, of course, can and must be supported; it is difficult to argue with the idea that governance should be ordered and predictable rather than chaotic and capricious. However, this general ideal runs into trouble when it is recommended as a realisable goal for judicial action: it has done as much harm as good. Because the ambition of *a government of laws, not persons* is both legally impractical and politically dubious, both the traditional defence and critical assault on the rule of law have created as many problems as they solve. Accordingly, a fresh appraisal and affirmation of the rule of law are required which recognise that democratic government has to blend both the contribution of *persons* and *laws*, such that justice and law are more often in harmony than in conflict.

Unfortunately, in jurisprudential debate, the old ideal-or-ideology problematic tends to take up far too much space and energy. Jurists line themselves up on one side or the other and then offer entrenched arguments about why the rule of law is good or bad in terms of its relation to a broader vision of social justice. Although presented as diametrically opposed in orientation and origin, the traditional and critical approaches are each the flip-side of the same formalist coin. In contrast to these rather jaded stances, I want to offer an account that takes a far less dogmatic view. The challenge is to throw away that devalued coinage and establish a more viable jurisprudential currency. Highlighting the contingent quality of all concepts and their practical merit, I maintain that the rule of

law can be both good and/or bad depending on both its agreed scope and its informing context. I want to propose an approach that emphasises that *laws* and *persons* interact in mutually-sustaining ways in the democratic practice of just governance; judges are neither so overwhelmed by laws such that their room for personal judgement disappears nor so untouched by the pull of laws such that their room for personal judgement is unconstrained. Accordingly, the more pertinent and pressing question is whether, in light of the history and its continued appropriation by liberal theorists, the rule of law can be salvaged and be given a radical reinterpretation that affirms its value to contemporary political theory and practice, while, at the same time, resisting the siren-song of liberal legalism. In an important sense, this essay takes up Peggy Radin's invitation to affirm the social significance and political relevance of the rule of law without also remaining committed to a traditional account of rules and rule-following, no matter how sophisticated or progressive. I offer a tentative answer to her question—"How can we deny formalism and affirm the rule of law?"[1]

This essay offers a minimalist or stripped-down account of the rule of law that is intended to present a very different account of democracy, adjudication and legislative responsibility from both traditionalists and critics alike. It will redefine not only the relationship between law and politics, but also what counts as law and as politics in a democracy. Nevertheless, although I believe that such a reaffirmation of the rule of law's importance is possible, I remain committed to the opinion that the rule of law has too often been used to stymie and frustrate popular participation and progressive action. However, I must also concede that, although the rule of law has been used to sustain elitist politics, it has also occasionally proved to be an effective principle to check the indulgent abuse of power by the few over the many. The fact that the rule of law has been used for reactionary purposes does not mean that it has to be or that it could not be used for more progressive ones. Accordingly, this essay is not a root-and-branch rejection of my earlier views, but an effort to give them a more pragmatic and less dogmatic spin. Indeed, I still hold to the conclusion that "there is a distinction between constitutional safeguards which constrain democratic activity in the name of democracy and those which constrain democratic activity in the name of 'right answers' ".[2]

This essay consists of five major parts. The first section explores the theoretical and practical questions to which the rule of law is supposed to be the answer. After an account of traditional and critical approaches, the second section suggests a different agenda of questions to be asked and answered. In the third section, reference is made to contemporary constitutional developments in Canada as a convenient site through which to draw out and initiate a more compelling account of a stripped-down rule of law. In particular, I focus on the recent decision of the Supreme Court of Canada on the legality of Quebec's

[1] Radin, "Reconsidering The Rule of Law" (1989) 69 *B.U.L. Rev*. 781 at 812.

[2] Hutchinson and Monahan, "Democracy and The Rule of Law" in A. Hutchinson and P. Monahan (eds.), *The Rule of Law: Ideal or Ideology* (Toronto: Carswells, 1987), p. 122.

possible right to secede unilaterally from Canada: the beauty of the judgment is that it treats the rule of law as only one principle in the constitutional compact and avoids the temptation to make it do more work than it reasonably should. The fourth section offers a general critique of rules and rule-following in law and legal theory from a non-foundationalist perspective. Introducing the role that "good faith" can play in adjudicative discipline, I sketch a minimalist account of the rule of law and its impact on constitutional interpretation. In the fifth section, I trace the implications of my non-foundationalist critique for the theory and practice of the rule of law. In particular, I respond to some of the likely objections and understandable reservations about my discussion and its institutional recommendations. Throughout the essay, there is a genuine effort to avoid taking sides in the continuing jurisprudential debate and, instead, to assess the merit of both courts and legislatures in terms of their contingent capacity to advance the contested cause of democracy.

I. THICK AND THIN

The familiar slogan of *a government of laws, not persons* is considered to be at the heart of the enduring attachment to the rule of law. This jurisprudential ideal is intended to highlight the need to guard against tyranny by ensuring that everyone, especially government officials, is subject to a pre-existing and public set of rules that can be applied in a reasonably objective and impartial way. Indeed, adherence to the rule of law idea is part of a larger sociological commitment which insists that, without the possibility of rule-ordered behaviour, society will lapse into chaos and arbitrariness.[3] Understood in this way, the rule of law is considered to be a vital component of a democratic polity. While it might not alone guarantee social justice, its strictures will help to ensure that public officials are kept in check and made accountable to citizens in their exercise of inevitable discretionary authority. In recent years, while jurists have begun to adopt a less severe version of what the rule of law might demand and be expected to do, there still exists the very real belief that it is possible to act and be guided by this institutional ideal. In particular, jurists and judges maintain that there is and should be a well-policed distinction between the legitimate judicial performance of rule-application and an illegitimate judicial exercise of rule-creation. Indeed, the attempt to define and defend that distinction has been at the heart of the contemporary debate over the rule of law in democratic societies. There are three main approaches to the rule of law—two traditional ones, a thick and thin variety, and a critical one.

The traditional approaches to the rule of law tend to coalesce around the central claim that rules can and should rule—rules are the basic currency of legal

[3] For a good intellectual history , see F. Dallmayr, "Hermeneutics and The Rule of Law" (1990) 11 *Cardozo L. Rev.* 1449 at 1451–9 and for a more jurisprudential account, see A. Hutchinson, *Waiting for Coraf: A Critique of Law and Rights* (Toronto: University of Toronto Press, 1995), pp. 7–15.

transactions, they have a core meaning, and such meanings should be relied upon to resolve most situations. The "thin" version of this traditional claim amounts to a constitutional principle of legality. Satisfying people's presumed demand for clear and fixed rules, these jurists strive to maintain a sharp demarcation of judicial and legislative authority; judges are bound by the law's commands as much as any other private citizen or public official. This is not so much an amoral stance as a moral position that defends a legalism of strictly rule-bound adjudication as the most morally-defensible account of law and adjudication in a constitutional democracy. It is a vision of judging that celebrates the systemic virtues of regularity, predictability and certainty over the concern with substantive justice in particular instances: formal rules are the most efficacious and legitimate way to protect substantive values. In its contemporary incarnation, it argues that, while the rule of law can best be understood as about the application of rules, it only demands a degree of conformity with past decisions: it is possible for a legal system to comply with the rule of law and still be undemocratic and/or unjust in general (i.e., apartheid South Africa) and in particular instances.[4] Because the adoption of such a view does not guarantee justice, many traditional jurists maintain that it must be supplemented with more substantive values.

The "thick" version insists that, while rules and their objective and impartial application are a vital part of any plausible account of the rule of law, law consists of more than rules. It holds that the existence of pre-announced, objectively-knowable and impartially-applied rules must be supplemented by tying such formal virtues to a substantive account of democratic justice. For such jurists, behind and within the rules is a political morality that guides and constrains judges when the application of rules is unclear or undesirable. Law is about values and politics, but not in any idiosyncratic or ideological way. In carrying out this jurisprudential manouevre, the primary task of theorists and judges is to detect and cultivate the politico-moral principles that breathe life into the dry husks of legal rules. In constitutional matters, the rule of law demands that positive law embody a particular vision of social justice that gains its constitutional justification and political appeal from the judicial enforcement of individual rights in the institutional exchanges between the state and its citizens. Accordingly, under such an account, judges are authorised to deal with political values so long as they do so in a neutral and objective way: "law . . . is deeply and thoroughly political . . ., but not a matter of personal or partisan politics".[5]

In spite of obvious differences, these traditionalists are united in their enduring formalistic belief that "rules can rule" and that there must be a clear and

[4] C. Sunstein, *Legal Reasoning and Political Conflict* (New York: Oxford University Press, 1996), PP. 190–5. In one of its most uncompromising incarnations, Justice Scalia (of the United States Supreme Court and former jurist) insists that "there are times when even a bad rule is better than no rule at all": A. Scalia, "The Rule of Law as a Law of Rules" (1989) 56 *U. Chi. L. Rev.* 1175 at 1179.

[5] R. Dworkin, *A Matter of Principle* (Cambridge, Mass.: Harvard University Press, 1984), p. 146.

defensible line between valid adjudication and ideological disputation. Those of both a thick and thin variety maintain that legal reasoning is a sufficiently detached and determinate enterprise which is capable of generating correct and predictable answers to social disputes in a way that marks it off, in a non-trivial and meaningful sense, from open-ended political wrangling. While immersed in politics and history, law is claimed to be its own thing and not entirely reducible to anything else. Without such a possibility, the fear is that the rule of law will be subverted and democratic governance will succumb to the tyranny of special-interest groups or partisan theories of social justice. Moreover, without adequate determinacy in legal discourse, judicial arbitrariness will become the order of the day and adjudication will collapse into a series of ad hoc and unprincipled encounters; "muddling through" will be all that is left. Accordingly, traditionalists all agree that the preservation of the rule of law "has the value of promising to make politics safe, of preventing Leviathan from becoming Frankenstein's monster, . . . [of imposing] real restraints on arbitrariness or despotic conduct".[6]

Needless to say, not all jurists have been convinced by the basic claim that rule-based adjudication can (to a greater or lesser extent) place "restraints on arbitrariness or despotic conduct". Some critics concede that, while the ideal of *a government of laws, not persons* has obvious appeal, they deny that rules can and do rule. Arguing instead that the rule of law more often acts as an ideological cover, they insist that rules can never impose sufficient restraints on judges and that adjudication will always be an exercise in arbitrariness or despotism; adjudication is more about reason in the service of power than power in the service of reason. On such a critical account, rules count for next to nothing in the fulfilment of judicial responsibility. In adjudication, rules only provide "a variety of rationalisations that a judge may freely chose from" and "the ultimate basis for a decision is a social and political judgment".[7] Indeed, in my earlier work, I was a vocal proponent of such a critical view. Decrying the judicialisation of politics as a travesty of the democratic ideal, I contended that the rule of law had been used as a constitutional barrier between governmental power and popular sovereignty and that it actually inhibited the flourishing of a rigorous democracy:

> "The Rule of Law is a sham; the esoteric and convoluted nature of legal doctrine is an accommodating screen to obscure its indeterminacy and the inescapable element of

[6] N. MacCormick, "The Ethics of Legalism" (1989) 2 *Ratio Juris* 184 at 188. Other prominent fearmongers are L. Fuller, *The Morality of Law* (New Haven: Yale University Press, 1969), p. 39; O. Fiss, "Objectivity and Interpretation" (1982) 34 *Stan. L. Rev.* 739 at 749; Edwards, "The Judicial Function and The Elusive Goal of Principled Decision-making" [1991] *Wisc. L. Rev.* 837 at 838–41; and Scalia, n. 4 above, at 1182.

[7] D. Kairys, "Politics and Law" (1984) 52 *Geo. Wash. L. Rev.* 243 at 244, 247, 245 and 247 and "Introduction" in *The Politics of Law: A Progressive Critique* 3rd edn. (New York: Basic Books, 1997), pp. 1–15.

judicial choice. Traditional lawyering is a clumsy and repetitive series of bootstrap arguments and legal discourse is only a stylized version of political discourse".[8]

The problem with the traditional and critical accounts of the rule of law is that they both operate within the same and very rigid notion of what "rules rule" can and might mean. The important and neglected question for both traditionalists and critics alike is not whether law in large part can be *represented* as the neutral application of objective rules, but whether it ever can *be* so. As regards adjudicative expectations, the root of the difficulty is the either/or understanding of the rule of law as representing a series of stark choices—between objective rule-application or subjective fiat; stability or chaos; authority or anarchy; and justice or oppression. A more nuanced and less crass appreciation of the role and possibilities of the rule of law in modern society is urgently required. At bottom, many of the critics of the rule of law (including myself) allowed themselves to be caught in a polarised debate in which rules either did or did not enable objective adjudication; there was a little space for a more sophisticated and less dichotomised position. Traditionalists strive to complete the foundationalist project of demonstrating that legal rules and their adjudicative application are grounded in something less contingent and more reliable than the shifting justificatory routines of present judicial incumbents. Although unalterably opposed to the worth or viability of this formalist project, the critics remain in thrall to its all-or-nothing character; they insist that legal rules and their adjudicative application amount to only expedient window-dressing for ideological manipulation. However, before presenting a more plausible account of what "rules rule" does and can mean, it is important to step back and examine the broader context of democracy within which the rule of law functions and which gives rise to the particular insistence that adjudication must be objective and impartial. It is my contention that both the traditionalists and critical accounts draw upon an impoverished concept and practice of democracy which leads them astray in their understanding of the role that courts might and could play in promoting its goals.

II. OF BICKEL AND PICKLES

In the sprawling debate over the democratic legitimacy of judicial review, the so-called counter-majoritarian difficulty has both energised and enervated constitutional law and theory. Given jurisprudential articulation by Alexander Bickel, it concerns the exercise of power possessed by judges, neither placed in office by the majority nor directly accountable to the majority, to invalidate majoritarian policies—How can a non-elective judiciary be justified in a

[8] A. Hutchinson, *Dwelling on the Threshold: Critical Essays on Modern Legal Thought* (Toronto: Carswells, 1989), p. 40.

democratic regime?[9] It should come as no surprise that jurisprudence has had much trouble with making any cogent response to the counter-majoritarian difficulty: it is based on loaded assumptions and misleading premises. While this theoretical and practical challenge puts appropriate emphasis on the tension between the enactments of duly-elected bodies and the actions of unelected ones, the assumptions that have been imputed to this inquiry and the answers it has prompted have tended to narrow and skew the debate in unhelpful ways. In particular, there are two basic themes which operate and which must be resisted—one is about courts and the other is about legislatures. These two ideas combine to offer a very impoverished notion of "democracy" and, in consequence, ensure that jurisprudential efforts to provide a viable and convincing account of legitimate adjudication are doomed to failure. The long and short of it is that the persistent attachment to Bickel has put both contemporary law and legal theory in something of a jurisprudential pickle.

The first underlying premise of the Bickelian problematic is that the politicised decisions of legislatures are democratic and in need of no further justification by simple virtue of the fact that they are the product of an elected assembly. Democracy is represented as essentially a procedural ideal: political decisions are validated through the ballot box and acquire political legitimacy in light of that process. Although more tacitly implied than overtly stated, there also seems to be the acceptance that legislatures are free to make decisions along the most partisan and ideological lines; reasoned decisions are an optional extra in this model of democratic governance. Indeed, it is acknowledged that legislatures do actually tend to act in unprincipled and opportunistic ways, with little genuine discussion and reflection about the larger issues of social justice for their own sake. As such, the legislative process concerns itself with the strategic aggregation of policy preferences as expressed through electoral procedures; principled deliberation over the values that might advance the cause of democracy are seen as window-dressing. Constitutional propriety is often only one more lever in an ideological set-to.

The second underlying premise of the Bickelian problematic follows from the first. In the same way that legislative decisions are presumptively democratic simply because they are the product of an elected assembly, so judicial opinions overruling legislative decisions are presumptively undemocratic because they are made by unelected officials. The fact that legislatures are considered unprincipled and beyond the constitutional pale adds greater urgency to the need to develop an adequate and defensible account of judicial review. Against such a backdrop, courts are always cast as the villain in the democratic piece. Expected

[9] A. Bickel, *The Least Dangerous Branch: The Supreme Court at the Bar of Politics* 2nd edn. (New Haven: Yale University Press, 1986), p. 16. See, also, Laurence H. Tribe, *American Constitutional Law* 2nd edn. (New York: Foundation Press, 1988), pp. 10–12; P. Monahan, *Politics and the Constitution: The Charter, Federalism and the Supreme Court of Canada* (Toronto: Carswells, 1987); and D. Beatty, *Constitutional Law in Theory and Practice* (Toronto: University of Toronto Press, 1995).

to perform in a way that is consistent with majority rule and not to engage in any competing imposition of values and preferences, the Supreme Court is "a deviant institution in [a democratic society"].[10] Accordingly, the vast majority of jurists have sought to develop a response to Bickel's counter-majoritarian difficulty that demonstrates how a non-elective judiciary can make decisions in a non-political way. In short, it has been claimed that such a challenge can be met by adhering to the demands and constraints of the rule of law—that rules rule—so that judges act in a principled and reasoned manner. While decisions by any non-elected body are always inferior, they might be able to attain some minimal legitimacy by scrupulous avoidance of any resort to political or ideological values which are the sole preserve of elected bodies.

Taken together, these two underlying premises combine to stymie and compromise any further or better elaboration of the rule of law. While there is a counter-majoritarian difficulty, it is neither the one that Bickel perceived nor is it the one which has dominated jurisprudence and constitutional theory. The focus on the counter-majoritarian difficulty is paralysing at all levels and results in too much leeway being given to legislatures and too little to courts. Or, more accurately, it demands that courts dissemble at best or lie at worst about what it is that they are doing. Nevertheless, the ideal of the rule of law persists because it gives support both to the power structure's legitimacy and people's need to believe that the status quo is, if not entirely just, then at least trying to be just.[11] Moreover, it is still widely maintained that any demonstration that adjudication is performed in a political and partial manner is a flaw that can be rectified by better laws or by better judges who understand the value and necessity of impartial and objective adjudication. In jurisprudential terms, the project of the rule of law remains valid despite its present invalid implementation. However, when the counter-majoritarian difficulty is given a richer and more compelling rendition, the possibility of providing a persuasive and practical account of the courts' role in a democracy is greatly enhanced.

As a way of illustrating and beginning that task, I will examine a recent decision of the Supreme Court of Canada and attempt to utilise it as a launching-pad for a more honest and adventurous account of the rule of law. While the Supreme Court has made considerable headway in moving away from the Bickelian paradigm and restating the jurisprudential challenge in terms of the democratic tension between constitutionalism and majoritarianism, it has still managed to retain the same old ambitions and ideas; it cannot or will not abandon the sustaining belief that this balance can be fixed in some enduring way and that the role of the courts in policing this balance can be effected in an impartial and politically-neutral way.

[10] Bickel, n. 9 above, at p. 18.
[11] Hasnas, "The Myth of The Rule of Law" [1995] *Wis. L. Rev.* 199.

III. A DEMOCRATIC REFERENCE

Perhaps more than most, Canada is a country that has a continuing debate about its constitutional arrangements. This debate covers not only the legal structure of such arrangements, but also the process by which such a structure can connect to the political debate for its alteration. Although this leads to more than its fair share of national angst, Canada has at least been obliged to attend to the legitimacy and substance of the basic building-blocks in its constitutional tool-kit. Of course, at the heart of this contemporary debate is the persistent problem about French-speaking Quebec's continued relationship with the rest of Canada. This takes many different shapes and forms, but the pressing issue is under what conditions if any can Quebec determine its own constitutional and political fate. This brings to the fore a whole host of difficult and enduring concepts and practices—democracy, sovereignty, self-determination, federalism and, of course, the rule of law. In the past year, this debate and issue has taken centre stage in Canada's constitutional drama, with the Supreme Court of Canada being required to provide its legal judgment on whether and under what circumstances Quebec might be able to secede unilaterally from Canada. The decision of the Supreme Court is an object lesson in the dilemmas that confront any theoretical efforts to give meaningful and legitimate practical content to the rule of law in a modern constitutional democracy. For present purposes, it is the nature of the problem that it outlines rather than its proffered solution that ought to command jurisprudential attention. The fact is that the Supreme Court seems to have offered a more sophisticated account of the problem and its possible solutions than much jurisprudential reflection.

The main question to be answered was "under the Constitution of Canada, can the National Assembly, legislature or government of Quebec effect the secession of Quebec from Canada unilaterally?" (para. 2)[12] The Supreme Court decided that it could not; any political decision to secede is constrained by and must be implemented in accordance with existing constitutional commitments. However, in a subtle analysis of the relation between democracy and the rule of law, it also held that, if there was a clear democratic vote in favour of secession, the rest of Canada would be obliged to negotiate with Quebec over the terms of its withdrawal from the Canadian union. Balancing constitutional rights and obligations as well as legal structures and political initiatives, the Supreme Court sought to clarify the delicate interplay between law and politics in a democracy and its own role in that dynamic confrontation. For instance, it decided that, whereas the legal order of the constitution prevented unilateral acts and required collective action, what constitutes "a clear democratic vote"

[12] Re Quebec Reference (1998), 161 DLR (4th) 385. References to this judgment are henceforth made in parentheses in the text by reference to paragraph numbers in the Court's official report. There is much that could be said about the strategic quality of the judgment—its unanimous nature; the timing of its release; its even-handedness, etc. I will not pursue these matters here.

and "legitimate negotiations" was a political matter that fell outside the legal mandate of the courts. As a unanimous Supreme Court concluded, "the task of the Court has been to clarify the legal framework within which political decisions are to be taken 'under the Constitution' and not to usurp the prerogatives of the political forces that operate within that framework" (para. 153). In reaching its specific decision and justifying it generally, the Supreme Court made three judicial moves that resonate strongly with the themes of this essay—the importance and limits of democratic process; the nature and status of the rule of law; and the relationship and balance between the two.

Recognising that the written constitutional rules must be interpreted in light of the underlying unwritten principles that have been developed over time, the Supreme Court took the view that any particular ruling must incorporate both sources of constitutional law: the enacted text is to be understood against the foundational principles of democracy, federalism, the rule of law, and respect for minority rights. Despite formalists' protestations to the contrary, it was emphasised throughout that constitutional texts are primary, but they do not exhaust the constitution and there is "an historical lineage" whose underlying principles "inform and sustain the constitutional text" (para. 49). One of the essential interpretive considerations is the principle of democracy. However, the Supreme Court realises that the meaning and demands of that principle are far from self-evident or universally accepted. While Anglo-Canadian constitutional history has tended to equate this with majority rule, democracy consists of much more. It is not simply concerned with the process of government: there is a substantive dimension that cannot be overlooked. According to the Supreme Court, these substantive goals include "to name but a few, respect for the inherent dignity of the human person, commitment to social justice and equality, accommodation of a wide variety of beliefs, respect for cultural and group identity, and faith in social and political institutions which enhance the participation of individuals and groups in society" (para. 64). However (and of vital importance to a non-foundational critic like myself), the Supreme Court concedes that what those values are, how they can be defined and how they interact is itself never fixed, but are part of the continuing debate over what democratic commitment entails: "a democratic system of government is committed to considering . . . dissenting voices and seeking to acknowledge and address those voices in the laws by which all in the community must live" (para. 68).[13]

In adopting such an approach, the Supreme Court seems to have made a giant leap beyond many of their juristic counterparts. It is surely the case that the assumption that legislatures have a lock on democratic legitimacy is seriously

[13] It should be clear that I strongly approve of the Supreme Court's insistence that such matters as equality, liberty and the like be treated separately from the rule of law. There are enough problems grappling with a thin version of the rule of law, without adding the unnecessary larding of the thick version. For a recent attempt to provide a thicker and more substantive account of the rule of law in Canada, see P. Hughes, "The Rule of Law and the Meaning of Equality" (unpublished manuscript, November 1998).

flawed. First, legislative outcomes are not majoritarian: legislative outcomes do not truly manifest majoritarian will and consequently their later setting-aside by courts, whatever it might do, does not upset decisions made by a majority. Individual legislators rarely claim to vote in line with the preferences of their constituents and, when they function as a group, there is no evidence to suggest that the outcome on any particular issue will coincide with majoritarianism.[14] Secondly, democracy's requirements are not exhausted by the establishment of democratic procedures. Most significantly, majoritarianism does not provide any substantial protection against legislative action against minorities; democracy demands more, therefore, than the facilitation of majority rule over minority interests. Thirdly, legislative action is not only about ideology, but ought to be about facilitating the representation of all citizens; there is no pressure under existing theories for legislatures to rise above the lowest standards and expectations so that politics can fulfil a more noble image of itself. Consequently, under any Bickelian-inspired account, legislatures are left free and clear in their (dis)regard for substantive values of democracy. A different account (like the one at which the Supreme Court is hinting) might have the virtue of pressing legislatures to take more seriously the need to address the justness of their actions against a broader and more substantive account of democracy.

Whereas a majoritarian account is premised on the mistaken claim that democracy is only about process, a fuller account of democracy involves much more; procedure is important, but it has no inherent or enduring superiority in achieving results that are democratic. Hence the notion and practice of a *constitutional democracy*. Democracy must mean more than simple majority rule or self-government, but must have a substantive element that both justifies the power of duly elected government and, at the same time, limits what can be done in the name of collective self-government. In particular, democracy must mean more than majoritarianism because this can result in many people being denied self-government. In short, democracy has both a substantive as well as procedural dimension.[15] It is hard to suggest that a constitutional arrangement can warrant the description "democratic" unless it possesses both a certain minimum content (e.g., no slavery) and a minimum process (e.g., popular elections). No matter how ideal the content or process might be, a society cannot claim to be democratic without some mix of the two. Democracy, therefore, is about social relations as well as political procedures; the one feeds off and reinforces the other. Democratic procedures cannot remain democratic unless they are utilised by people who share some basic equality and liberty. Similarly, a society that comprises people who relate on equal and free terms cannot be democratic unless it allows and incorporates certain procedures for popular government.

This broader and more nuanced understanding of democracy, of course, leads to an obvious difficulty—the two fundamental principles to which consti-

[14] Farber and Frickey, "The Jurisprudence of Public Choice" (1987) 63 *Texas L. Rev*. 873.
[15] See F. Michelman, "Brennan and Democracy" (1998) 86 *Cal. L. Rev*. 399 at 421.

tutional democracy is committed are, at worst, incompatible and, at best, in the most severe tension. According to one principle, the will of the citizens as expressed through the available political procedures should govern and any limits on this exercise of popular power are unjustifiable. But this principle competes with another equally important principle. This holds that the majority cannot do whatever it likes in the name of democracy; there are certain outcomes that cannot be tolerated in a society that claims to be just, no matter how democratic the procedures that gave rise to them.[16] For societies to be worthy of the label "democratic", there must be a balance between the procedural and substantive dimensions which, being contingent and contextual, will change and vary over time.

In order to operationalise this view of democracy as demanding more than majority rule, the Supreme Court recognises that popular sovereignty has to be supplemented and constrained by other constitutional principles. Along with federalism, a major principle of the Canadian constitutional order is the rule of law. After acknowledging that it is "a highly textured expression" (para. 70) that lends itself to diverse interpretations, the Supreme Court confirms that the rule of law is a basic requirement of any stable, predictable and ordered society. Staying at the relative safety of the most abstract level, the Supreme Court identified three major components to the rule of law—the existence of one constitution and set of rules for both government and private persons; the creation and maintenance of an actual order of positive laws which embodies the more general principle of normative order; and the insistence that the exercise of all public power, including that by the courts, must find its ultimate source in a legal rule. Accordingly, explicitly adopting a thin version of this constitutional axiom, it maintained that "the rule of law principle requires that all governmental action must comply with the law, including the Constitution" (para. 72).

For the Supreme Court, therefore, majority rule combines with other constitutional principles, such as the rule of law, to ensure that democracy is implemented and respected in a procedural as well as substantive way:

> "The consent of the governed is a value that is basic to our understanding of a free and democratic society. Yet democracy in any real sense of the word cannot exist without the rule of law. It is the law that creates the framework within which the 'sovereign will' is to be ascertained and implemented. To be accorded legitimacy, democratic institutions must rest, ultimately, on a legal foundation. That is, they must allow for the participation of, and accountability to, the people, through public institutions created under the Constitution. Equally, however, a system of government cannot survive through adherence to the law alone. A political system must also possess legitimacy, and in our political culture, that requires an interaction between the rule of law and the democratic principle. The system must be capable of reflecting the aspirations of the people . . .

[16] S. Holmes, "Precommitment and the Paradox of Democracy" in J. Elster and R. Slagstad (eds.), *Constitutionalism and Democracy* (New York: Cambridge University Press, 1988), pp. 196–7.

> Constitutional government is necessarily predicated on the idea that the political representatives of the people of a province have the capacity and the power to commit the province to be bound into the future by the constitutional rules being adopted. These rules are 'binding' not in the sense of frustrating the will of a majority of a province, but as defining the majority which must be consulted in order to alter the fundamental balances of political power (including the spheres of autonomy guaranteed by the principle of federalism), individual rights, and minority rights in our society. Of course, those constitutional rules are themselves amenable to amendment, but only through a process of negotiation which ensures that there is an opportunity for the constitutionally defined rights of all the parties to be respected and reconciled. In this way, our belief in democracy may be harmonized with our belief in constitutionalism. Constitutional amendment often requires some form of substantial consensus precisely because the content of the underlying principles of our Constitution demand it. By requiring broad support in the form of an 'enhanced majority' to achieve constitutional change, the Constitution ensures that minority interests must be addressed before proposed changes which would affect them may be enacted. It might be objected, then, that constitutionalism is therefore incompatible with democratic government. This would be an erroneous view. Constitutionalism facilitates—indeed, makes possible—a democratic political system by creating an orderly framework within which people may make political decisions. Viewed correctly, constitutionalism and the rule of law are not in conflict with democracy; rather, they are essential to it. Without that relationship, the political will upon which democratic decisions are taken would itself be undermined (paras 67 and 76–8)."

For the Supreme Court, therefore, majority rule is not tantamount to democracy and does not take precedence over all other values and principles in the Canadian constitutional order; any other argument misunderstands profoundly the meaning of popular sovereignty and the nature of a constitutional democracy. In reaching this understanding, the Supreme Court clearly assumes that such ideals are attainable in the sense that "rules rule"; its approach is premised on the claim that rules can be stated and applied in a relatively determinate and uncontroversial way in the vast number of circumstances. This continued preoccupation with the need to ground an objective practice of judicial interpretation that obviates judicial value-choice and that does not tread on the democratic toes of legislative decision-making is doomed to failure: the history of twentieth century jurisprudential and constitutional theory has been dominated by almost endless futile attempts to provide an account of the courts' role that is consistent with the democratic priority of majoritarianism. As judicial review involves unelected judges invalidating the actions of elected legislators, all judicial review is anti-majoritarian and, therefore, presumptively undemocratic; no theory can reconcile judicial review with majority rule.[17] And, perhaps more importantly, this continued search for the jurisprudential grail is unnecessary. Having abandoned the crude Bickelian challenge to the democratic legitimacy of courts, the Supreme Court should follow through on the political logic

[17] E. Chemerinsky, *Interpreting the Constitution* (New York: Basic Books, 1988), pp. 11–12 and Chemerinsky, "Foreword: The Vanishing Constitution" (1989) 103 *Harv. L. Rev.* 43.

of its own analysis; it must have the institutional courage of its own jurisprudential convictions.

Liberated from the confining strictures of the Bickelian problematic, the question of how and whether courts act with democratic legitimacy is of a very different order and character. The Bickelian counter-majoritarian difficulty has little to say about what values are important to democracy other than an unthinking regard for majoritarian processes. Once the principle of democracy is accepted to have a substantive as well as formal dimension, the justification for judicial action must also be viewed in substantive as well as formal terms. The work of courts need not be judged by their capacity to be objective and impartial nor by their willingness to be consistent with and not interfere with majority politics. Instead, they can be evaluated in terms of the value choices that they make and the contribution that their decisions make to the promotion of democracy in the here-and-now. If the Bickelian assumptions—that legislatures are unprincipled and political and that courts are principled and reasoned—are dropped, it is possible to arrive at a very different understanding and account of the relation between courts and legislatures. For instance, the conclusion is possible that legislatures and courts are both principled and unprincipled to greater and lesser extents at different times and that each can further (as well as inhibit) the cause of democratic justice on a particular issue as well as the other. The more pressing conundrum, therefore, is that, if democratic procedures do not guarantee democratic outcomes and democratic outcomes need not result from democratic procedures, how can we best organise constitutional arrangements so that democracy as a whole is more than less likely to prevail. Accordingly, the appropriate inquiry in a constitutional democracy is not to ask whether the courts have acted politically and, therefore, improperly, but whether the political choices that they have made serve democracy. Moreover, in a democracy, what counts as "democratic" is contingent and contextual. Because this is substantive and rhetorical, not formal and analytical, it will always be a contested and contestable issue—law is politics.

Nevertheless, what counts as a democratic decision is not entirely reducible to a political and, therefore, open-ended debate about what is most appropriately democratic at the time and under the circumstances. The formal dimension of democracy insists that some account is taken of the general institutional location and position of relative governmental agencies: the fact that legislators are elected and judges are unelected has some political salience. This is where a stripped-down version of the rule of law has its part to play. The democratic bottom-line is that, unlike legislators, judges are expected to respect and observe the gravitational push and pull of established legal doctrine. However, in playing by the existing rules of the legal game, it has to be understood that the rules of the legal game are themselves in play because what it means to play the game is an always debatable and never closed matter. To put it more conventionally, the simple injunction for judges to "apply the rules" is never simple because what counts as a rule and what is involved in applying it cannot be placed

beyond contestation or dispute. Again, law is politics in the sense that what counts as law is never not a political matter. Accordingly, in the next section, I offer a non-foundationalist account of how "rules rule" that is intended to be distinguished from the formalist claims of traditional jurists and the nihilistic assertions of the critics.[18]

IV. HOW RULES RULE

The fact is that the ideal of an impersonal application of laws is unattainable: there can be no professional application of rules which does not involve the exercise of personal values. However, this acknowledgment does not presage an inexorable decline into palm-tree justice or arbitrary governance. In an important sense, the critics throw out the progressive baby with the traditional bathwater. I contend that the judges are both more and less constrained by rules than any of the critical or traditionalist critical jurists think. While rules do not rule, they are not merely so much chopped liver. Judges are more constrained in the sense that they cannot get completely outside of rules and exercise an entirely free choice, but they are less constrained in the sense that they are not obliged to reach any particular decision as the result of a commitment to resolve disputes through rule-application. It is not so much that rules do not exist, but that they do not exist as canonical directives whose meaning is available without interpretation and which can impersonally dispose of cases. The meaning of a rule and its application never simply *is*—it is something to be argued for or about and not something to be argued from. As conventional and social matters, the meaning and application of rules is contextual and, therefore, open: the fear of rule-free choice is as contrived as faith in rule-bound choice. Any claim that judges are only being held back from a frenzy of arbitrariness by rules' restraining power is not only unrealistic, but it does scant credit to the integrity and efforts of most judges. Consequently, I resist the temptation to deify or demonise the act of adjudication and, in the process, turn judges into demi-gods or demagogues. My account will not only serve to chastise those traditionalist who are hubristic enough to project law as a seamless web of doctrinal filaments, but will also act as a much-needed corrective to those critics who are driven to present rule-application as nothing more than a transparent exercise in self-delusion.

Traditionalist jurists are mistaken when they assert that any and all alternative accounts of law and adjudication "reject reason-giving all together, putting in its place power, or play, or conventions".[19] The stark contrast between a reluctant attachment to a foundationalist account of law and adjudication, with its promise of coherent, determinate and correct answers, and the capitulation

[18] For a full account and defence of this approach, see A. Hutchinson, *It's All in the Game: A Critical Account of Law, Politics and Adjudication* (Duke University Press, forthcoming).

[19] C. Sunstein, "Analogical Reasoning" (1992) 105 *Harv. L. Rev.* 741 at 779, n. 130.

to a faithless world in which arbitrary power and ideologues run amok is contrived. The kind of non-foundationalist critique that I offer does not deny the possibility of reason-giving or shared understanding. While reason and power are inseparable, they are not identical or entirely collapsible into each other. In the same way that reason cannot detach itself completely from power and determine the terms of its own cogency, power does not consume all of reason and reduce politics to only arbitrary and random power-plays. While my non-foundationalist approach does undermine the practice of legal reasoning as a complete and grounded craft, it does not depict judging and decision-making generally as only the quirky or idiosyncratic performance of the Nietzschean will to power: Zarathustra is not the non-foundationalist judicial alter-ego. In contrast, while a non-foundationalist account refuses to understand adjudication as separate from politics, it does not subscribe to an "anything goes" ethic. In so doing, it maintains that meaning-giving is not simply subjective and narcissistic, but is constrained in the same way that all individuals are constrained and constituted by the context of relations and conventions within which they find themselves. My non-foundationalist critique, therefore, is not beholden to an account of adjudication that comprehends it either as a foundationalist act of commitment to an abiding Reason or as a critical surrender to a capricious irrationality: adjudication is based on reason insofar as it is constructed in and through the very judicial arguments that it is intended to guide.

Consequently, the jurisprudential claim that I am actually making is that judges do not stand outside the rules, but inhabit the rules in a particular way. Judges (and jurists) are always situated within a context of freedom and restraint; they are neither never fully restrained nor ever entirely free. In law's language game, judges give meaning to rules and their own lives in the constant struggle to negotiate the forces of freedom and constraint that constitute the rules and their own lives; they are freely restrained and restrainedly free. Indeed, freedom and constraint can only be made sense of reciprocally. The rule-engaged judge is neither context-dependent nor context-transcendent; the rules and the rule-user are enmeshed in a mutually affecting relationship. While there will be occasions of insidious manipulation, it is as mistaken to suggest that the judge has complete interpretive freedom as it is to claim that the law exerts complete interpretive control. In the same way that people (re-)constitute themselves in their lived relations with others, so judges (re-)constitute themselves in their participation in law's language game. While there is nothing fixed or determinate about this dynamic interaction, judges will continue to change as they engage in the play of difference that constitutes the adjudicative practice. Indeed, the claim to have moved beyond or got outside the play of power is the most Machiavellian of power-plays. A non-foundational account of adjudication takes seriously both the restrictive pull of rules and the liberating push of interpretation. It is not possible to dispense with rules nor with the need to interpret them; each is what makes the other tick. There is not only no set of rules that stands apart from interpretive attempts at its hermeneutical appropriation,

but there is no literal interpretation—one that claims to be transparently representational rather than opaquely constitutive—against which other interpretations can be contrasted.

While I agree with the critics' argument that "legal reasoning is not a method or process that leads reasonable, competent and fair-minded people to particular results in particular cases" and that "the ultimate basis for a decision is a social and political judgment", I do not agree that law is only "a variety of rationalisations that a judge may freely chose from" and that "the decision is not based on, or determined by, legal reasoning".[20] While it is no doubt accurate that some judges do use legal reasoning as only a rationalisation for an existing prejudice, this characterisation of the judicial performance fails to incorporate the fact that most judges do base their decisions on legal reasoning in that they make a good faith effort to interpret and deploy legal rules as the argumentative resources and occasions for their decision. Any other view commit its advocates to the untenable proposition that all judges are engaged in a giant and unavoidable hoax when they claim that their decisions and judgments are based upon and restrained by the rules of legal doctrine. In contrast, I maintain that it is not so much that judges ignore the rules as that they could not follow the rules even if they were minded to do so—the availability of rules as heuristic directives for decision is simply denied by many critics. Moreover, any jurisprudential critique that insists that rules count for next to nothing in the fulfilment of judicial responsibility tends to exacerbate the formalist fear that, without some plausible account of determinate rule-following, there will be an official anarchy in which rules will count for nothing or simply be used as crude *ex post* rationalisations for *ex ante* decisions: political wolves will simply be clothed in judicial sheep's clothing.

As such, the requirement to act in good faith is at the heart of a revised understanding of the rule of law. Beginning the reasoning process with a more or less definite conclusion in mind is not the problem: most judges start with some more or less vague notion of where they think that their judgment should go or come out. The key issue is how they relate or connect that conclusion to the legal materials. A minimal democratic restraint asks that judges make some genuine effort to support that conclusion by reference to the rules; judges must work *within* as well as *with* law's rules and argumentative resources. This justificatory requirement of "good faith" is not unique to adjudicative practice. While it demands honesty and sincerity in one's actions, these qualities amount to only part of the requirement. To accept that one must act in good faith is to recognise that the end does not always justify the means and that sometimes things might not always go. Of course, there is no reason to assume that two people who act with integrity will agree on what is right and wrong or on a common course of action. Good faith cannot in itself tell people what to do in situations of uncertainty and doubt; it can only give them the courage to act on their own

[20] Kairys, n. 7 above, at 244, 247, 245 and 247.

convictions in a way that takes seriously the responsibility to act fairly.[21] Nor does it insist upon a stubborn consistency; people act with integrity when they are neither dogmatic nor unprincipled, but are open to engagement and, therefore, change. In short, the requirement to act in good faith is the difference between an attachment to the irresponsible claim that "anything goes" to the principled commitment that "anything might go (but also might not)". Accordingly, good faith can be thought of as acting in line with the spirit of the enterprise in which one is engaged and respecting other people's expectations about what is supposed to happen. It demands, at a minimum, that this should be done without furtively or opportunistically imposing one's own view of what should happen; it implies more than avoiding outright cheating and deception.[22] While this standard is open and indeterminate, it imposes a constraint on action, even if there are no objective or uncontroversial norms for compliance and even if the constraint of good faith is always in play. While there is a difference between those who do and do not adjudicate in good faith, there is none between those who play by the existing rules of the game and those who play to change them. Each judge is engaged in the political practice of adjudication; it is simply that they are making different choices about particular rules and their application in particular contexts.

For instance, in matters of constitutional interpretation, there is a significant difference between judges who are trying to make sense of a particular text, even if that process does generate multiple and contradictory meanings, and judges who are not making that effort at all. This is particularly the case in circumstances where a judge has made the effort to engage the text or rule, but chooses to ignore the results and, instead, knowingly grafts an extraneous and different meaning onto a particular text.[23] This constraint of good faith does not in any way limit or restrict the kind or range of meanings that can be attributed to different rules; the judge is still to free construct meaning within the enabling and constraining context of the rules. However, whatever interpretation is offered or whatever application is suggested, it must result from a genuine effort to make sense of the rule in hand or to deploy law's argumentative resources in a

[21] S. Carter, *Integrity* (New York: Basic Books, 1996), p. 7.

[22] I offer no account of the psychological basis of "bad faith"—how judges explain to themselves what they are doing when they are judging, especially in response to the charge that they are acting ideologically rather than legally, is not my concern. See D. Kennedy, *A Critique of Adjudication* (Cambridge, Mass.: Harvard University Press, 1997), pp. 180–212. In arguing that "adjudication is best described as ideological choice . . . carried on by actors many of whom are in bad faith", he includes any judge who existentially denies that adjudication is so practised; see ibid. at p. 4 and J-P. Sartre, *Being and Nothingness* (H. Barnes (trans.), New York: Washington Square Press, 1965), pp. 89–90. Kennedy is addressing much more than those judges who cheat deliberately and wilfully by following their ideological predisposition without an attempt to fit it to the available legal materials. I tend more toward the view that "nobody would claim that their own thinking was ideological just as nobody would habitually refer to themselves as Fatso . . . Ideology like halitosis is in this sense what the other person has": T. Eagleton, *Ideology: An Introduction* (London: Verso Books, 1991), p. 3.

[23] See S. Fish, "Play of Surfaces: Theory and The Law" in G. Leyh (ed.), *Legal Hermeneutics: History, Theory, and Practice* (Berkeley: University of California Press, 1992), p. 297 at p. 302.

conscientious way. Understood in this way, the requirement of good faith is more an issue of moral integrity than a matter of analytical accuracy; it is less about legal rightness and more about political reasonableness. Whether a particular judicial performance is accepted or rejected by others will not be because it is somehow validated by reference to a foundationalist algorithm as a correct or true interpretation of the rule. A judgment will gain acceptance or rejection because the judge is able to persuade others that it is a reasonable interpretation and that it has earned its legitimacy through the persuasive force of its supporting arguments. To argue that only good renditions of particular performances can count as valid performances would mean that a "poor" performance would have to be treated as oxymoronic. Within conventional practice, it is quite sensible to insist that a bad performance or even parody of an artistic work is no less a performance of that work.[24] As such, the distinction between legitimate and illegitimate behaviour is not a bright line, but a spectrum of options and choices. It is not possible to withdraw to some grounded position outside the contested area to validate or vouchsafe objective standards for evaluation or criticism: there is no escape from the responsibility of choosing among competing values. To ask more is to flirt with the seductive appeal of a foundationalist approach, but to ask less is to succumb to the false allure of the critical approach.

Under my non-foundationalist account, it might well be possible for someone somewhere at sometime to formulate rules or apply familiar rules in what presently appear to be bizarre or perverse ways. However, in order to count as a good faith act of legal judgment, all that must be demonstrated is that such judges hold a practical and actual belief that the rules do permit such a course of action. If there is such a felt sense of what a rule or rules might plausibly be interpreted to mean, the debate over meaning will be joined as one about substantive wisdom as opposed to interpretive correctness. Such judges might be persuaded that there are better or more compelling ways to apply a particular rule, but that debate will be a rhetorical engagement over substantive reasonableness, not a philosophical reflection over formal truth. If such judges are not persuaded by the arguments of their colleagues or commentators, the upshot will be that their ensuing judgment will simply fall by the doctrinal wayside as one more failed effort to convince the legal community of the judgment's political merit or good sense; it ought not to be that such judges are condemned as having failed to fulfil their institutional role properly or that their judgments are rejected as invalid. Adjudication not only allows for and accommodates bad performances, but it also actively encourages adventurous performances which are intended to transform its operating rules and which will risk failure. The fact that judges arrive at different or even contradictory applications of particular rules is less of a concern than the failure of judges to take the rules seriously. While judges are obligated to apply rules, this does not oblige them to reach cer-

[24] See J. Passmore, *Serious Art* (London: Duckworths, 1991), p. 65.

tain decisions. Of course, to concede that there may be several applications of a rule that count as valid or legitimate is to confirm that law is rife with indeterminacy and that adjudication is inescapably permeated with political choices. As such, traditional jurists need not so much fear that "anything goes" as a matter of judicial decision—that judges might decide cases on the basis of a cointoss, the litigants' hair-colour, astrological charts or the like. It is that, as a matter of rule-application and through a process of rational disagreement, judges might justify almost any outcome—"anything *might* go". Again, such a non-foundational account does not reduce adjudication to brute preference nor dispense with the possibility of rational debate. It simply insists that the idea of rational debate be understood in a way that recognises that "rational debate" is part of the debate and not a grounding for its continuing performance or critique.

V. CURIAL IMPLICATIONS

My proposal is centred on the rule of law as a principle of institutional morality that imposes constraints on the exercise of official authority and discretion; it seeks to stabilise, not to strait-jacket, the exercise of such discretionary authority. It places a check on the activities of judges such that they are not free to do whatever it is that they wish without some genuine and serious effort to follow the principles, procedures and practices that are in play, including and especially those principles, procedures and practices that cover how changes in the principles, procedures and practices can be made. Admittedly, the constraint of "good faith" that is suggested is very limited and is, of course, itself indeterminate and always in play.[25] Nevertheless, such a thin account of the rule of law is the price that has to be paid for a constitutional device that is both theoretically viable (i.e., it plays a suitably modest and complementary role in the overall scheme of things) and practically workable (i.e., it asks judges to do what is actually possible rather than what might wishfully be thought desirable). Like the general mix between the rule of law and its other constitutional components, the democratic demands of the rule of law will be context-specific: sometimes, it will play a minor role and sometimes, it will play a much larger one. For example, in societies, like Eastern Europe and Latin-America, that are in transition from repressive rule to democratic governance, what is considered just and appropriate will depend upon the contingent extent and intransigence of prior injustice. The rule of law may have a valuable role to play in facilitating that

[25] While I present an unabashedly thin version of the rule of law, I am not suggesting that it is politically neutral as some proponents do. See, for example, R. Summers, "A Formal Theory of the Rule of Law" (1993) 6 *Ratio Juris* 127. I justify my particular account of the rule of law in reference to a particular vision of democracy that is itself never beyond political debate or defence. As such, my account is not intended to finesse the political status of the rule of law, but to confirm the unavoidable political character of its idea and practice.

shift and, "rather than grounding legal order, it serves to mediate the normative shift in justice that characterises these extraordinary periods".[26] Indeed, it might be that, in certain circumstances of crisis and upheaval, the judiciary are temporarily best placed to effect large-scale changes in a potent and telling manner.

By combining the arguments about the revitalised role of courts in a democracy and the political nature of rule-application, the rule of law begins to assume a more viable and less alienating presence. Indeed, such a revised formulation has a number of important implications for legal theory and practice. To begin with, it strongly suggests that legal theory is not in the game of governing or predicting legal practice: to posit such a purpose is to misconceive law as an inorganic clump of prescriptive rules and to mischaracterise adjudication as a technological exercise in normative analysis. It is mistaken and impractical to persist in believing that the role of legal theory is to ground a finite set of fixed rules which result from and inform legal practice, but which are somehow apart from legal practice and not entirely reducible to it. The tendency to view law as being the methodological measure of the rational and the task of judges as being the formulaic application of that metric is the very antithesis of what I maintain that judges do, can do and should do. Because rules are indeterminate, there can be no rules about the application of rules that do not themselves fall back on the need for judgement. Rather than engage, as almost all contemporary mainstream jurisprudence does, in "a futile striving to overcome the essential unruliness of judgement",[27] legal theory has a much more modest task—uncovering assumptions, illuminating contradictions and making contextual suggestions for law's contingent improvement. Jurists are not the grand architects of law; they are more its humble odd-jobbers.

When adjudication is understood as a practical activity rather than a theoretical reflection, judges are liberated from the perceived need to pronounce *ex cathedra* by putting into practice currently-favoured theories of adjudication. Instead, they are expected to do nothing more (and nothing less) than engage with the general legal materials at hand and fashion them so as to fit them to the specific requirements of the issue before them in a way that advances justice through law: it is the legal past that impels them forward, not pulls them back, in their yearning for justice. As such, great judges do not ignore the past or obsess about it, but work the past so as to realise its present possibilities for future innovation. Whether this satisfies the rationalistic fantasies of contemporary jurisprudence is almost besides the point: judges are to be judged by the political merit of their practical performances, not the conceptual coherence of their theoretical reflections. Further, from my non-foundationalist perspective, inquiries about whether law and political morality are necessarily connected and about whether judges' resort to political morality is impersonally determined are no longer pertinent or pressing. By engaging in law and adjudication,

[26] R. Teitel, "Transitional Jurisprudence: The Role of Law in Political Transformation" (1997) 106 *Yale L.J.* 2009 at 2016.

[27] P. Nonet, "Judgment" (1995) 48 *Vand. L. Rev.* 987 at 989.

its professional participants are involved with political morality in a way that is neither impersonal nor determined; judges and jurists can no more choose to ignore or finesse issues of political morality than they can slough off their responsibility for actually resolving those issues by implicating law's inherent logic. There are no "hard cases" in that all cases are both hard and easy in proportion to the (in)stability in social judgements and historical contexts that energise the dynamic interaction of determinacy and indeterminacy in legal language; there is no way to apply the rules to the facts that does not involve ideological choice between competing values and contrasting interpretations. There can be few better examples of the insight that "the formal rules of constitutional texts can remain the same, but their meaning and application can change" than Canada's constitutional history. The enacted constitutional structure governing federalism has not changed in any substantial sense over the 140 years of Canada's existence; the same wording is meditated upon today as it was in 1867. However, as even a nodding acquaintance with Canadian constitutional history confirms, constitutional doctrine has evolved and changed in response to political and social circumstances.[28]

However, the fact that law is deeply and thoroughly political as a matter of personal and partisan politics is not cause for regret. Rather than resist its normative force, the challenge is to get beyond apocalyptic despair or fatalistic resignation and, instead, to turn it to democratic advantage. This can be done by recognising that the insight that "law is politics" is an opportunity to enhance social justice, not an obstacle to its achievement. In making ideological choices, judges best meet their democratic responsibilities neither by masking their political commitments nor by grinding a favoured political axe. Instead, they must be prepared to address their values candidly and self-critically. Indeed, judges ought to treat their political convictions in much the same way that they treat legal materials—as resources to be interrogated and re-worked in the service of a democratic vision of social justice that is itself always in the process of revision and transformation. The hallmark of the great judge is not to carve out a detailed theory of justice and to cling to it through doctrinal thick and thin; this is to mistake political hubris for democratic duty. Instead, great judges recognise that, while they must assume the responsibility of political choice, they do this in a suitably democratic manner by "keeping questions open, lingering in enlightened confusion, so that [they] do not miss the next insight when it comes".[29] In this way, judges will be able to square their substantive political commitments with their formal democratic responsibilities: they will recognise that, at times, the spirit of the rule of law demands that they place particularised justice ahead of systemic consistency.

Furthermore, looked at from the perspective of a revised rule of law, the perennial puzzle over how to make a theoretically justifiable and practically

[28] See P. Hogg, *Constitutional Law of Canada* 3rd edn. (Toronto: Carswells, 1995).
[29] Meyer, "Is Practical Reason Mindless?" (1998) 86 *Geo. L.J.* 647 at 652.

sustainable distinction between the work of judges and legislators in a democratic society begins to take on a very different complexion. While contemporary jurisprudential scholarship is beginning to move beyond the debilitating effects of the Bickelian paradigm, there remains a distinct reluctance to abandon the felt need to allocate institutional responsibility in accordance with an enduring and detailed plan. A non-foundationalist approach recognises that, while the tension between constitutionalism and majoritarianism is enduring, there is no enduring or constant way to resolve that tension. Efforts to identify and defend a suitable distinction are doomed because there is none.[30] Or, at least, there is no distinction between adjudicative and legislative practices that can be maintained in a way that is itself foundational or fixed; any such distinctions or practices do not ground or fix debate because they are always part of the debate to be had. The Supreme Court's pronouncement that "our democratic institutions necessarily accommodate a continuous process of discussion and reflection" (para. 150) has two important implications when read from a non-foundationalist perspective. The first and more obvious one is that what democracy demands is itself always open to "discussion and reflection" and, of course, change: what counts as democratic is a contingent and contextual calculation. The second and overlooked one is that the court is one of those "democratic institutions". In contrast to the Bickelian depiction of courts as "a deviant institution",[31] it is much more accurate and appealing to understood the courts as democratic institutions which have a vital and complementary role to play in the "continuous process of discussion and reflection" about what democracy means and demands. It is less that courts and legislatures are in competition or conflict, but that they each have an important and political role to play in defining and defending the cause of democracy.

Of course, there have been and will continue to be occasions on which that conflict is marked and on which courts pursue a very different set of substantive values to those of the legislatures. However, the existence of disagreement need not always be a fearful sign of democratic crisis; it might simply represent a productive exchange over democracy. Crisis is as likely to be precipitated by clinging to the Bickelian belief that legislatures have the corner on what is and is not democratic and that courts must necessarily be subservient to such decisions. As I have tried to suggest throughout this essay, this mistaken conception of the relation between courts and legislatures does more harm than good; the debate about democracy is carried on in veiled and oblique terms. However, my alternative non-foundationalist approach avoids such rash and blunt evaluations by insisting that both courts and legislatures are in the same game, namely fashioning and implementing a notion of democracy that can provide practical answers to the challenges that presently confront society. And, in doing that, neither courts nor legislatures has a lock on political judgement about what it is

[30] For the leading attempt to do this, see R. Dworkin, "The Partnership Conception of Democracy" (1998) 86 *Cal. L. Rev.* 453 at 458.

[31] Bickel, n. 9 above, at p. 18.

best to do. In a thriving and progressive democracy, it is the substance of the results reached as much as the procedural site of their making that should merit critical attention. It is a travesty of that democratic aspiration to allow the broader debate over constitutional democracy to be reduced only to a question of institutional legitimacy.

Within the non-foundational approach that I adopt, the presumption that decisions of legislatures or governments are necessarily more democratic than those of courts is flawed; courts are "not uniquely subject to the influences of regressive social pressures and the risk of strategies being ineffective or backfiring".[32] The respective responsibilities of judges and legislators cannot be defined outside of the never-ending debate about what democracy demands and what best serves democracy in a particular historical setting and socio-political context. In the same way that judges are not constrained in their decisions by legal doctrine in a final or objective sense such that "anything might go", so legislators are not free to make decisions that are unencumbered by institutional constraints such that "anything goes". In both cases, the practices of adjudication and legislation only make sense within a contextual set of legal rules, institutional expectations, interpretive constraints, material conditions, social views, intellectual circumstances, etc. that both make any decisions possible and also influence the kind of decisions that are made. While it is tempting to suggest that a less-than-ideal (elected) legislative body is preferable to an ideal-as-possible (unelected) judicial bench, this kind of generalised and once-and-for-all assessment is to be resisted. Of course, all commentators and critics will approach specific problems with their own prejudices and presumptions, but they must work to put them in political and judicial play. In particular, they must avoid the foundationalist tendency to allow generalisable principle to be the enemy of pragmatic good and must resist the critical temptation to permit the political end to justify the institutional means.

The dogmatic commitments of both traditional and critical jurists have forced them into unnecessarily entrenched positions on particular judicial decisions and institutional arrangements. Whereas traditionalists feel obliged to defend most decisions of the courts, critics tend to feel an equal compulsion to condemn almost all decisions.[33] Furthermore, the lack of nuance and room for manoeuvre in traditional and critical accounts of the rule of law means that they can often find themselves in intellectual solidarity with an unappealing collection of political colleagues. Indeed, the critical account of the rule of law in particular has played into the hands of right-wing popularist groups. For instance, in Canada, the right-wing Reform Party has deployed the sweeping rhetoric of critical jurisprudence to challenge the activism of the Supreme Court in dealing

[32] J. Bakan, *Just Words: Constitutional Rights and Social Wrongs* (Toronto: University of Toronto Press, 1997), p. 147.

[33] Contrast Hogg, n. 28 above, and M. Mandel, *The Charter of Rights and the Legalisation of Politics*, 2nd edn. (Toronto: J. Lorimer, 1997).

with, among other matters, gay rights and aboriginal land claims.[34] In a very smart move, the Reform Party has been able to hide its particular homophobic and racist agenda in a more general and less offensive critique of court activism. Adopting the Bickelian logic of jurisprudential critics, it has framed its opposition to the adoption of certain substantive values in terms of the formal constitutional roles of Parliament and the courts:

> "Respecting the proper roles of Parliament and the courts is crucial to the health of a democracy. Duly elected Parliamentarians have the unique responsibility to debate, write and amend the nation's laws. Courts can interpret them in light of the Constitution and declare them valid or invalid, but should refrain from creating public policy or rewriting statutes in their decisions. Courts should properly defer to the elected representatives of the people rather than charging ahead with their own ideas about the wording of Canada's laws".

This posturing is less a debate about democracy and more one about the wisdom of certain substantive decisions. While this, of course, could be cast as a debate about democracy, so could any dispute over any particular decision whose substantive basis was controversial. Within the impoverished jurisprudential terms of the debate over judicial review in a constitutional democracy, the Reform Party gain unwarranted legitimacy for political views that are directly contrary to a richer account of democracy's just requirements. In sketching the thrust of the rule of law, the Supreme Court stresses that it "provides a shield for individuals from arbitrary state action" (para. 70). This is not a chastisement against judicial activism and, still less, a prohibition against the resort to political values in judicial decision-making. While it should be read as a caution and a reminder to judges about their institutional location in the relative constitutional scheme of things, it simply places an injunction on them not to act in a capricious or irresponsible way. By acting in good faith, they more than satisfy this general constraint. The blind adherence to precedent and principle is more likely to lead to arbitrary action than a reasoned and reasonable judicial effort to advance the cause of democratic governance. Indeed, as represented in the Bickelian approach, judges have a hard time acting in a way that is not arbitrary.

In contrast, I have suggested that democracy cannot be reduced to a purely procedural ideal and that evaluations of the courts' performance cannot be framed only in terms of the activist/passivist paradigm. When a more substantive account of democracy is accepted, assessments of the courts' contribution will necessarily be less sweeping and more particularised. Moreover, the critical slogan that "we may grant until we are blue in the face that legislatures aren't wholly democratic, but that isn't going to make courts more democratic than legislatures" is deprived of its disarming purchase.[35] It is not so much that courts

[34] See *Delgamuukw v British Columbia* [1997] 3 SCR 1010 and *Vriend v Alberta* [1998] 1 SCR 493.

[35] J. Ely, *Democracy and Distrust* (Cambridge, Mass: Harvard University Press, 1980), p. 67.

are more democratic: there is nothing in my arguments that commit me to defending courts generally or any decisions particularly that they make. However, the contrary is also true—nothing commits me to condemning courts generally or any particular decisions that they make. Whether specific decisions are good or bad will be a contextual and contingent decision in light of a situated judgement about what democracy demands and what it requires at that particular time. In the same way that what counts as democratic must always be up for grabs in a democratic society, so which institution is better placed to act effectively in accordance with democratic values will always be up for grabs.

Much is made in the Supreme Court's judgment about the distinction between law and politics. The Supreme Court was at pains to emphasise that "the task of the Court has been to clarify the legal framework within which political decisions are to be taken 'under the Constitution', not to usurp the prerogatives of the political forces that operate within that framework" (para. 153). While this is an important issue for courts and commentators alike, it is an entirely political matter—the question of what does and does not amount to law or politics is a deeply and pervasively political inquiry. After all, the Supreme Court made it plain that it was "the court's *own* assessment of its proper role in the constitutional framework of our democratic form of government" (para. 26) that was being offered and that was decisive. Moreover, as with any political assessment, the distinction will shift and change in its legal application and political implications. For the Supreme Court to announce, therefore, that some matters are political and out of its jurisdiction does not in any way confound the claim that "law is politics". On the contrary, when understood from a non-foundationalist perspective, it confirms that the decision as to what is legitimately considered law (and, therefore, appropriate for judicial consideration) and what is political (and, therefore, not) is a highly political matter whose resolution will be contingent, contextual and contested. As the informing political vision changes, so will the distinction between what is and is not politics. Moreover, any judicial decision to designate something as "political" and better suited to the domain of legislative politics—the issue, for example, of what constitutes a "clear majority", a "clear question", and "good faith negotiations"—does not mean that everything else is not political and can be legitimately dealt with by the courts. It is more useful to understand the Supreme Court as dividing responsibility among democratic institutions, including courts, for handling political matters. It is not that what the courts are doing is any less political than what other official agencies of the state are doing, only that such political matters are better suited to judicial treatment than legislative consideration, at least at the present time and under existing circumstances.

Finally, some commentators have begun to argue that the rule of law is not only a fixed ideal in the Canadian constitutional firmament, but that is also has a fixed meaning in that it is determinate in the scope and substance of its expectations. For instance, in an imaginative essay, Patrick Monahan contends that, on the basis of a growing body of judicial pronouncements, the rule of law has

acquired a constitutional status in Canada that would authorise courts to invalidate legislation without any express constitutional provision that relates directly to the matter in hand (i.e., property or contract rights). Indeed, he goes so far as to suggest that the rule of law "represents a substantive limitation of Parliament's ability to pass laws".[36] I resist such a foundationalist conclusion. I am not claiming that the rule of law cannot be interpreted to stretch as far as Monahan wants; this is a matter of political persuasion, not accurate legal analysis. However, I resist entirely Monahan's implicit claim that his expansive interpretation is the best, necessary or only interpretation of this constitutional principle. Insofar as he claims it be a descriptive statement about what the law is, it seems a highly implausible and unconvincing account. Furthermore, insofar as he claims it to be a prescriptive statement about what the law ought to be, the imposition of such a sweeping constraint on legislatures is by no means the best, necessary or only interpretation of what best serves democracy in present Canadian circumstances. While Monahan's interpretation is radical in the best sense (i.e., it is creative and paradigm-shifting), it is also conservative in the worst sense (i.e., gives enormous weight to past decisions and their intractability).

Moreover, as the Supreme Court hints in the *Quebec Reference*, what the rule of law does and does not demand in particular situations and how it interacts with other constitutional imperatives is not fixed or inflexible; "these defining principles function in symbiosis: no single principle can be defined in isolation from the others, nor does any one principle trump or exclude the operation of any other" (para. 49). Accordingly, the rule of law, like the Supreme Court's interpretation of it, is never entirely one thing or the other in any final or permanent sense: it must be judged by their contingent resonance with prevailing circumstances and contextual conditions. Moreover, because "the democracy principle . . . cannot be invoked to trump . . . the rule of law" (para. 91), I presume that rule of law cannot trump the democracy principle: both principles must be accommodated as optimally as possible. Of course, on occasion, one will be given greater contingent weight than the other, but this is very different than a constitutional hierarchy of principles in which one principle has absolute priority over another. So, in some circumstances the rule of law has outweighed individual rights, whereas in others it has been counter-balanced by other constitutional consideration.[37] It is important to stress that while the same ingredi-

[36] P. Monahan, "Is the Pearson Airport Legislation Unconstitutional?: The Rule of Law as a Limit on Contract Repudiation by Government" (1995) 33 *Osgoode Hall L.J.* 411 at 416. The courts' reliance on the rule of law as a separate constitutional principle, begun in *Roncarelli v Duplessis* [1959] SCR 121, has continued apace in recent years. While it is reinforced by the Charter's preamble that "Canada is founded upon principles that recognise the supremacy of God and the rule of law", it has been treated by the courts as having independent constitutional status. See, for example, *Patriation Reference* [1981] 1 SCR 753; *Re Manitoba Language Rights* [1985] 1 SCR 721; *OPSEU v Ontario (A.G.)* [1987] 2 SCR 2; and *Re Remuneration of Judges* [1997] 3 SCR 3.

[37] See the *Manitoba Language Rights* n. 36 above (more important to preserve the legal order than it was to give immediate effect to a constitutional declaration of invalidity).

ents prevail, it is their mix and balance that will always be in issue and beyond final fixing. And, of course, there is no meta-principle that can determine what the correct mix is—that is part of the continuing debate about democracy itself. To re-phrase the favoured imagery of constitutional discourse, it is less that the constitution is "a living tree" and more that it is a living forest, a political ecosystem that comprises symbiotic parts that shift and transform in their relationships as the social and economic circumstances change.[38]

VI. CONCLUSION

In the mid-1980s, Owen Fiss published an influential essay that lamented the critical turn in jurisprudence. Capturing mainstream jurists' discomfort and disappointment, he regretted that many critical jurists had begun to turn their backs on the law and its adjudicative potential to contribute to the articulation and implementation of a more egalitarian vision of society. For him, the failure of the courts to live up to their transformative promise was cause to redouble efforts at jurisprudential revitalisation, not to abandon them. In particular, he chastised critical scholars for trivialising and endangering "the proudest and noblest ambitions of the law" to function, through its curial sites, as an official arena through which to forge and sustain a public morality. Fiss's challenge and complaint are well taken. However, contrary to his depiction of critical jurisprudence, I do not offer a critical account of law and adjudication that is distinguished by "the purity of [its] negativism". While some critical accounts of law and adjudication might be vulnerable to such chastisement, my non-foundationalist account is neither purely negative nor negatively pure; a ludic understanding of adjudication incorporates both the constructive and deconstructive dimensions of law and also emphasises the practical and contextual quality of adjudicative practice. Indeed, I agree with Fiss that "what is required is that judges be constrained in their judgment, and that they certainly are".[39] My disagreement is with the nature and force of those constraints.

Judges are both free and constrained in their efforts to fulfil their institutional responsibilities: they do not stand outside the law, but inhabit the law in a particular way. Being always situated within a legal context of freedom and restraint, judges (and jurists) are never fully restrained nor ever entirely free. The law-engaged judge is neither context-dependent nor context-transcendent, but is situated within a mutually affecting relationship with the law. While there is nothing fixed or determinate about this dynamic interaction, judges will continue to change as they engage in the changing play of difference that constitutes law and its adjudicative practice. In adopting such a playful account of law that takes seriously both the rules' restrictive pull and their interpretation's

[38] See *Edwards v Attorney-General for Canada* [1930] AC 123 at 136 per Lord Sankey (PC).
[39] O. Fiss, "The Death of Law?" (1990) 72 *Cornell L. Rev*. 1 at 1, 9 and 11.

liberating push, I am not committed to a jurisprudence in which adjudication will be, as Fiss fears, *only* particularistic or exclusively instrumental. Instead, as well as being both particularistic and instrumental, a non-foundationalist account of the rule of law is able to explain and cultivate a mode of adjudicative performance that best captures its sense as a peculiarly professional practice (in which it stands as something of its own thing) and as a profoundly ideological undertaking (in which it is organically related to the larger context of society): it accepts the foundationalist claim that adjudication is a sincere effort to engage with legal materials and, at the same time, holds on to the anti-foundationalist claim that law is an exercise in political choice. At a minimum, the rule of law demands that, as well as combining "sceptical commitment . . . with . . . an ironic self-awareness of the contingency of one's own beliefs . . . [and] tolerance for contradiction", judges and jurists should proceed *as if* constitutional and legal argument mattered.[40]

The upshot, therefore, of treating adjudication as a playful and experimental performance in law's language game will not be, as Fiss concludes, "law without inspiration". Nor will it be "the death of law . . . as we have come to admire it". A non-foundationalist account of the rule of law is both much less pessimistic and much more realistic than the naively optimistic Fiss concludes. While my critical approach might well lead to the "death of law as we have known it throughout history", it can contribute significantly to the rebirth of law as "a generative force of our public life" and the revival of jurisprudence as an "inspiration" to judicial practitioners.[41] A non-foundationalist account of law, politics and adjudication does not deliver a public morality, but it does offer a public practice through which to regenerate a morality that respects the playful quality of law and life. Moreover, in a society that still holds to the present institutional arrangements, the courts might (and, of course, might not) become a site for a series of political and transformative practices that people can not only admire, but which also might serve to inspire them. Mindful that "anything might go", the value of any judicial or juristic contribution is always "up for grabs"; the prospect of it being good is always shadowed by the risk of it being bad. Most importantly, law's adjudicative performances can never transcend the moral qualities and ideological visions of the individuals that are entrusted with this opportunity and responsibility. At the end of the day, "law is politics" and the need for a diverse group of political players must not be forgotten; democracy and particularly the rule of law demand no less.

[40] L. Seidman and M. Tushnet, *Remnants of Belief: Contemporary Constitutional Issues* (New York: Oxford University Press, 1996), pp. 200–1.

[41] Fiss, n. 39 above, at 15 and 16.

Part III

The Limits of Legal Order

11

Parks, Dogs, and the Rule of Law: Post-communist Reflections

ANDRÁS SAJÓ

In the following pages excrement appears many times, in many places. Because excrement is everywhere—and not circumlocution. You cannot step into circumlocution; it doesn't produce that repulsion that is ingrained in humans from childhood and that animal loathing that exists in dogs in the way that excrement does.

Furthermore, in the following pages you cannot always know whether I am talking about the dog or the owner. I do not have a reason to exonerate either one; I do not see the difference between them. I do not believe in the inherent worthiness of dogs, though newborn puppies are cute. I especially like bulldog puppies. I do not make a distinction between the dog defecating on the sidewalk and his owner doing the same. Although perhaps human excrement smells worse. I also do not distinguish between owners. In Budapest I have more than once seen an owner walking behind his dog with a plastic bag. He who picks up the mess is a moral being in the Kantian Sense. After all, if his deed's maxim would become common, there would be order. Yet in the midst of disintegration the Kantian ethics is inadequate. Especially in matters concerning dogs. The man with the plastic bag—objectively—belongs to the world of those with dogs. If he comes out against them—that's different. But in the spiral of disorder the Kantian ethic is self-deception. Kant is done for here and that's the problem. Because what kind of civil society can this be, if Kant is not valid—if he, in fact, does harm? Without Kant, of course, one slips back to being partisans of law and order. Order is usually the order of the stronger. The stronger ones, however, are the dogs. At least the stronger dogs.

What follows is the law and order of dog freedom.

Since birth I have lived beside a park. The park is about as wide as an apartment building, but so long, that we, the children who lived at the upper end, almost never ventured to the lower end. In the middle of the park a large pool was built, so we were able to bathe in summer. There was a park warden as well, with a club, until the end of the 1950s. The park deteriorated each year, but only slightly. Here they closed the urinal; there they carried away the iron fencing.

But—according to my embellished memory—until the end of communism the park suffered only minor damage. At the beginning of Stalinism, under cover of night, they removed a park statue that was considered politically hostile; nevertheless, for long after it was still possible to jump from the mysterious pedestal which remained. In the 1970s they ruined the park with improvements in accordance with the spirit of the era: they erected several Communist Party statues and planted some Soviet trees which were not suitable for the park (for example, the birch). For the sake of economy they reduced the pool to half its previous size, but for a number of years there was still water in it. Few people allowed their children to swim in it after the dogs started to go there to cool off.

Maybe it was at the same time as this that the dogs appeared in the sandbox too, coming up against violent indignation. The authorities put up a sign prohibiting the dogs and though they continued to come, they did not dare to remove it. There was communism, but with a human face, not a dog's jowls. There were plenty of dogs, but on leashes. At least during the day they were on leashes. Only in the late 1980s did they take their dominion. By the beginning of the 1990s the new generation, in groups of three or four, ran around without leashes while the owners, dressed in their comfortable and practical leisure clothing, enjoying the warmth of their new found community, exchanged their pithy ideas, especially—I suppose—about the beauty of having a dog. The dogs preferred the grassy, spacious, open area to the walkway—in agreement with their ball- and broken branch-throwing owners. In this way they took dominion over the lawn as well, that is, over that which according to the landscape company that sowed it in spring would have been grass, and which up until then small children had tried to trample down while playing soccer with their fathers.

Before long the dogs and their owners remained alone in the park, at least during the dog hours. The curfew was lifted daily: for a few hours love-sick students and adulterous wage-earners went there to kiss, grandmothers sat in the sun with their grandchildren. Between the two dog-walking parts of the day pensioners playing cards on two-forint boards sat on grey concrete benches and lounged around grey concrete tables placed there in some sudden burst of the people's democracy's magnanimity. The unemployed, staggering as they came from the downtown streets, armed with beer and cheap pocket sized brandy bottles, watched them play. Now and then somebody, perhaps the drunk bums (or tramps, or more politically correct, the homeless) sheltering themselves here during the night turned the trash cans upside down. The dogs indifferently walked around a thick wrapping paper left from a half a pound piece of bologna.

By 1997, for some mysterious reason, the authorities decided to renovate the then sixty-year-old park. From the very outset they did not plan to rebuild the playground. The dog-owning engineer who studied the site experienced no small children in the park. If they aren't there now, then why make a playground? The engineer was right. The children of dog-frightened parents disappeared, just like the joggers.

But the accusation that the park is full of filth is not at all true, because the dogs also relieve themselves on the streets leading to the park, making dark streaks on the house walls and sidewalks, as well as on the wheels of the cars along the sidewalks. Between rainfalls, the walls get darker and darker. Because more and more dogs arrive, from farther and farther away, from the streets of the has-beens, to drag with themselves even those who are only just getting ready to slide down.

Or it has rained more and more rarely.

The dogs came, emancipated, unleashed, without muzzles, without a disciplinary word.

People without dogs were silent. The dogs' owners were also silent. Everyone walked around the dog excrement and the parked cars pressed against the sidewalk.

And no one says a word.

Since the collapse of communism Budapest is one of the most dynamically developing cities in dog-keeping. The growth—like the development of mini-capitalism in general—began even under "goulash communism", but it was only brought to completion during freedom. Consumer capitalism became production capitalism—on the dog front as well.

The performance of the Hungarian economy (as traditionally happens in Eastern economies) hides from all forms of transparency that runs the risk of paying taxes. We can therefore only estimate the growth in the canine sector. We can make our conclusions based on the ruined parks, the piles side-stepped during a walk of about ten minutes, from the multitude of dog food commercials, from the number of people torn apart by bulldogs.

Dog-keeping and its customs in Budapest provide the first solidifying outlines of a new social order after state socialism.

Both the rich and the poor post-communist, he who would like grow rich, as well as he who fears falling into poverty, he who is elegant, and he who cannot afford other company or entertainment, all have dogs, though, of course, different breeds and in different ways. There is a market for guard dogs, as well as for attack dogs, and fighting dogs are needed for competitions. Dogs overrun the concrete boxes of the apartment blocks, because there is a market for their progeny. On the edge of my park a laser-printed poster, with an address, advertises Rottweiler puppies for sale. The homeless as well as the wealthy get dogs; they both fear attack. Neo-Nazis in ostentatious uniforms march with fighting dogs all along the boulevard. The passers-by tremble as they read the traditional placards. The policemen likewise. They complain that they aren't allowed to have such dogs and they don't have any idea what they could, that is, what they should, do about the march. But the gypsies keep mad dogs too; they are expecting attacks from the Neo-Nazis. On Klauzal Square, the servants of the skin-head Capulets and the gypsy Montagues, trained to love blood, growl behind the urine-marked borders, while the trembling public health service employees stealthily throw civilisation's symbol, i.e. lime, on the walls, the medieval defence against contagion.

As they multiply, they become more and more conscious of themselves. The masters. The dogs simply follow their masters. The master's consciousness seems to be full of mental confusion. The masters are as they are. The confusion which only slumbers, in humans comes to light in the dog which signifies their achievement. Here is the message: they don't give a shit about me—heck, they don't give a shit about anything! The dog fulfils one's most beautiful dreams.

Every kind of deprivation, constraint, and boorishness sides with dog-keeping. Or perhaps the wage-earner simply wants to give some joy to his children, and to his misfortune, he cannnot provide the Hungarian paradise: to live in a house with a garden, like the child's girlfriend or a television soap-opera family. The motive is indifferent: even a good intention will produce shit.

It is not the soul that is deformed, it is the city. It is mercilessly anti-dog. Not planned for dogs, just as it was not planned for cars. This hostile environment must be made, if turning it into dog-loving is no longer possible, then at least dog-like.

This is the secret master plan. And the dogs, led by their masters, take over the city. The city and its dog-less inhabitants are incapable of stopping it. The inhabitants are potential traitors: we, too, would keep a dog that was after our own hearts. After a while we forget that we are supposed to be opposed to them. That dog excrement is not fate. That on the streets of Basel or New York it is not predestination or centuries-old civic tradition that disciplines the dog-owners, but severe and popular official terror. The good people of Budapest patiently suffer and walk around with their heads *hanging*. Two itsy-bitsy bulldogs lounge about in front of my grocery store. They are already not so very itsy-bitsy. We cross to the other side. After all, there is a sidewalk there, too. The customers come back later. We also learn and make note of the exact hours when the Rottweiler's huge dark shadow alights on the sidewalk. We take the children and the puppies out for their walks in conformity with the Rottweiler's habits. Besides, the authorities also know that the Rottweiler lives here. The authorities also use this street. They know the address, but they don't know what to do.

There is no chance of opposition. Until the end of the 1980s the humility and goodwill inspired by the police state and legal insecurity still coloured the owner's excuse: "the dog doesn't bite". In the time of political change they hurled the same excuse along with a purposefully defiant "Mind your Business!" Today the inquisitive accent has disappeared. Or they yell—aggressively or threateningly, depending on the dog: "You don't like him? Well I don't like your face either!" (People with German Shepherds and bigger dogs, they address you informally. There is democracy, that is, they don't give a shit about anyone, and even less about those who are below them.)

The dog owner is post-communism's free man. He who is free, thinks the dog owner, cannot be restricted. No one, not even a puppy can be restricted. Not even the weak old pensioner widow who has a dog, in so far as she is a dog owner. In this respect even she is a dignified being. You can count on the fact that if some bothersome, shoe-urinating, before-dawn-wakening mongrel is kicked aside, other dog owners will be in solidarity with him. They will

hurry to stand beside him in the common struggle for control over the park or apartment building.

The hopelessness of the anti-dog opposition theoretically follows from Mancur Olson's paradox about collective action. The dogs, who are the minority, gain victory over the majority and their diffuse interests. The dog-less are weaker than the owner of the German Shepherd. Not only weaker, they also are not sufficiently motivated. What can one do as an individual? Shoot the barking beasts? Pay a janitor to drive off the dog that urinates in the entranceway? Perhaps voluntarily pay more taxes for cleaning up the mess? In the end, contrary to the intention, what will happen is that the tax will be turned to the dog's good in a programme to establish dog-running paths.

The dog-less are not at all capable of organising themselves; there is no chance of a collective campaign. A few intellectuals write articles if they step into it more than once some day. Many isolated articles have appeared in recent years. Their only effect was that the trade journals, in the name of the dignified dog breeders, rejected the libels, the crude generalisations . . .

Those who suffer from the dogs, since they are incapable of organising any kind of special collective action, expect protection from the state, which is supposed to be a substitute for collective action.

In principle, democracy could also help, because there are more voters who have an interest in not stepping into excrement, or in their children or grandchildren using the playground, than there are savage dog owners. And yet no one has seen a candidate for municipal or national government who would promise dog cleanliness. Any mayor who would like to create order through fines and removal of the dogs had better dig his own grave with the dog-catcher. Money for a dog-catcher? Never. And for that you don't even need too many dog-friendly organisations. The creator of order would not come into conflict with the dog keepers' existing, otherwise excellent organisations. It is not even truly in the dog breeders' interest to protect the interests of the mongrels and the illegal breeders who operate in apartments rented from the municipality. They would prepare such a programme on a political basis, and in fact easily, because general disgust would greet it. The liberal and the conservative parties would both attack it: among us fines are not the solution; instead, they would give something from the public funds to the dog owners also, to make them feel good. The state should not discipline, but give away. Discipline is something that is extremely unpopular in a place where everyone trespasses in some way or another (e.g., parking). Today they go after the dogs, tomorrow they'll want to tow away my car for the trumped-up reason that you can't cross the sidewalk because of where it's parked. Who needs this?

He who needs law and order should protect himself. Let's not leave it to the authorities: experience tells us that the authorities need order so that at this address they can extort tax or bribe money from those who they have picked out at random and declared to be disorderly or well to do. A modern moral philo-

sopher (if such a profession still existed) would say that he who is accused is not guilty, just unlucky.

The maintenance of public order does not stand a chance. The state, the authorities, are not going to serve as a substitute for collective action that is impossible to establish. The post-communist state and its officials, beaten by the rule of law (just like its stupid citizens, our feeble-minded people forged from practical, talented individuals), declare Bankruptcy. If not the dogs', at least the dog owners' freedom prevails. Every freedom is the enslavement of others if it knows no bounds, if people aren't taught and trained to respect the limits, that is—speaking in the language of dog trainers—if they aren't trained to sit up and beg. The dog owners will be the oppressors. (Of each other as well, but let that be their problem).

Those who don't have dogs could tolerate the fact that others do if appropriate rules protect them—and if somebody actually abides by these rules. The observance of a rule, according to the sociology of law cliché, requires mobilised and controlled enforcement of the rule by the interested parties, for that however, an enforcing mechanism is needed. The smaller the opposition of those who are regulated, the more effective the enforcement. In our case, every condition is lacking.

A much-modified metropolitan decree, originating in 1968, regulates dog-keeping in Budapest (3/1968 (IX.30)). (Oh, holy legal continuity—or rather, even the communists were able to arrange dog matters on paper). In the capital, just as in other Hungarian cities, dog owners must have their dogs registered at a veterinarian and they have to vaccinate them at least against rabies. A dog can be in a public place, but it has to be on a leash and, if it is vicious, it has to be muzzled. Otherwise, vicious dogs must be tied up during the day. Whether or not the dog is vicious, is of course only determined *afterwards*, after a regretful accident, though who can examine a dog's past on the street? Since 1994 certain breeds of dogs, under any circumstances, have to be fitted with a muzzle and, theoretically, the most that could cause problems for the authorities is recognising the specified breeds. The dog may only defecate at marked places. There are sanctions as well, with hefty sums. In case of a violation of the rule about keeping a vicious dog locked up, after a further official written warning the dog must be taken away. With a local permit, dogs can be bred in an apartment building or tenement house; only the consent of the public sanitation authority is needed, as well as the consent of the *majority* of the neighbours. Please pay attention to the starting point! Freedom to practise dog-breeding in the apartment or tenement building is measured out as a discretionary act (is this an act of grace?).

Furthermore, again with an official permit, the requirements of which are unknown, you can also keep more than one dog in an apartment in Budapest. For this you do not even have to ask the neighbours, though the local self-government administrator, *out of public interest*, can prohibit dog-keeping in the building, after a hearing with the residents. (Legally, the inhabitants perhaps would not be completely defenceless; since the permit originates in administra-

tive proceedings, there they can have legal standing as a party in the case because it pertains to their legal interests.)

If the dog disturbs the residents' peace, "if the dog owner does not ensure the prescribed conditions for keeping the dog, the removal of the dog shall be ordered from the apartment, or in some other way the authority shall arrange for the cessation of the reasons for the complaint". Consequently, if somebody doesn't like the dog, he first of all has to commit himself to quarrelling with the dog owner, possibly even getting bitten by the dog living in the building, and subsequently after some maintenance conditions were established and violated, another complaint is needed, possibly from more residents, since the regulation uses the plural. One irritatedneighbour's irritation is not irritation.

The violation of any of the notification, leash, muzzle, or breeding and maintenance obligations carries a fine of up to 3,000 Hungarian forints, which can be increased to 10,000 forints in downtown Budapest. The fine can be as much as 20,000 for a public sanitation offence, which is a substantial sum, as high as the monthly subsistence wage. On the other hand, it is only possible to impound dogs that are kept in the neighbourhood of schools and playgrounds, or those *consistently* disturbing the peace or bothering the calm of the neighbours. It is not proper to remove the vicious, without muzzle, or regularly messy dog. What kind of rule of law state would it be, where for a dog's behaviour (for not putting on a muzzle) the owner can be punished with confiscation of his property?

These feeble regulations come with such a huge official fuss and with such dog-scorning, and perverse neighbour-reporting and burdens of proof, that from the outset there is no chance of them being used. The starting point of the regulation is pro-dog, that is, in principle you can keep a dog; a few pro forma inconveniences are added to it. Yet some smaller cities dare to be pro-resident and pro-home owner, perhaps because in a smaller community the democratic pressure is stronger. In the noble cities of Csongrád, Székesfehérvár and Balatonfüred you need the consent of *all of the residents* (tenants and owners) in order to keep a dog. Dogs may only walk around with a muzzle—all kinds of dogs, perhaps with the exception of lap-dogs. In Székesfehérvár it is forbidden to take a dog into a park even on a leash, while in Budapest and Balatonfüred a dog is only prevented from entering a park if the mayor or local self-government administrator has particularly forbidden it. And why incur additional work and disputes? In Balatonfüred you can only let your dog run in the outer districts, yet how could I burden the people of downtown Budapest with something like that? After all, there is equality before the law, and the downtown doesn't have any outer districts. Therefore, the administrator could designate an untrafficked area for letting dogs run without leashes on the basis of a proposal from the Dog Breeders' Association and with the agreement of the police, but without consulting the inhabitants. Dog exercise is a technical matter, and a dog's right. The practice is even simpler. The dog owners form a running path for themselves according to the demands of the dog, even out of the park—the most that their

association does is make the act official. Here we are provided with an application of the famous legal thesis, the normative strength of the *factual* prevails in the public domain.

But, after all, public sanitation, health, and safety rules (muzzles) exist, and they even carry serious fines. But where is enforcement? Who is going to establish the offence, who is going to collect the fine? Hardly the police, more important matters crop up for them. Right here is a criminal investigation. To ask for protection concerning dogs and barking today is to facilitate the escape of serious criminals. The Public Works department (which in the matter of the vicious dogs in the apartment building is of course not at all competent), recently appears to require self-defence spray, but it doesn't take the risk of action against dogs. It only likes to fine the law-abiding person. It can somehow deal with an abandoned vehicle, but with an abandoned dog? Even worse, with one who has an owner? And likewise, who would take the trouble to report illegal activity in the park? And how should the complainant prove it? After all, this is a rule of law state—whoever alleges must prove as well. How much wrangling would it involve if the local self-government administrator, in the spirit of the rule of law, would "effectuate the removal [of the dog] from the house"? As a matter of course he cannot conduct a search of the premises; what sort of a thing would it be to deprive the owner of the protection of his home just because he has a dog? Without a search warrant and actual search, however, it is impossible to verify that the dog is not kept according to the regulations. What is the proof to the complaint that the dog is kept in a way that "disturbs the peace"? Or let me be (as an exception) pro-dog: how can it be established that the owners violate the animal protection laws, that they torture the animal? What kind of rule of law state would it be where you can presume that a dog threatens with bodily harm? What if the dog growls and pulls at the leash? That just goes to show that it has a leash. That in the past it bit somebody? This does not mean anything for the future because the past, as Hungarian history shows, is not the pawn of the future, or if it is, then we should put it in a pawn shop instead of using it as an argument.

Legal security, due process, the presumption of innocence, freedom of property, perhaps even personality rights (at best not extended to dogs)—all of these are on the side of the dog owner. Of course the noble precepts also serve the civil servant's convenience—inaction. The rule of law values are on the side of the dog owner, not so much because after years of oppression we have become sensitive emotionally to rights and individual claims, but because in these years of confusion people do not know that words may have meanings. Or they do not want to know, because knowledge is uncomfortable. (The maintenance of ignorance naturally belongs to the cocksure logic of transformation, for example, it facilitates that stealing of public property that is called accumulation of capital or efficient ownership). Misunderstanding, most of all misunderstanding about freedom of property, and the behavioural order built on misunderstanding, is a characteristic of the post-communist era. Is it transition? And to what? Here every kind of property is ill-gotten in some way and has to function while being

dogged with troubles: maintaining the scam that surrounded the acquisition, never paying back the loan that was taken, concealing the fact that the legal title or the compensation voucher used as legal tender is false, or that oblivion awaits the obligations stated in the tender. The statute of limitations, the acquisition of wealth protected by trade secrets and the right of personality, and the sacrosanct freedom of contracts (even if it assures one-sided advantages); these are the rule of law state's holy trinity (or, better, *quartetto*) on which the post-communist system of acquisitive rule of law is founded.

Keeping a dog is an honourable thing. Or it would be, if the owner would assume all the costs for having his dog—if he would pick up the excrement and put cork on the walls of its room, so that the barking does not bother the neighbour. Or the dog owner who lives in a room facing the courtyard should acknowledge that he should buy a nice house with a garden beside the sheepdog. But in post-communism equality of opportunities extends only to the opportunity to have a dog.

The responsibility that goes along with property would be too great a burden. The owner does not take it upon himself. And the state does not take it upon itself to enforce the responsibility. The law is reasonable: it understands the cost of coercion—out of incompetence it saves its own incompetence as well.

The history of the law regarding the keeping of dangerous dogs demonstrates the misunderstanding of freedom and the paralysis facing it much more dramatically than does the above discussion. Freedom brought the fighting dogs to Hungary. Demand increased for guard and attack dogs, and of course for everything from which you can make money fast. Dog fights and pitbull breeding proved to be such a thing. Once the Red Army left there was an immediate leap in the number of injuries caused by dogs. In early 1994 the severely pro-order conservative government banned dog fights, but they did not dare to make public the original draft of the regulation which included much stricter restrictions. Following this, a few pedestrians were torn apart by pitbulls in 1994. Because of this, a question was addressed to the Minister of the Interior in Parliament, who said that in that year the police destroyed fifteen attack dogs. (In the following years the police, who considered themselves authorised to act only in order to protect someone's life, intervened with similar frequency. There was hardly ever enough evidence against suspect pitbulls; the neighbours' reports before and after the events are not worth anything in the view taken by a strict rule of law doctrine.) In 1996 a fighting dog tore apart a three-year-old little girl. From this umpteenth episode the opposition stirred up a political row in Parliament. The government, in order to alleviate the pressure, agreed to enact a strict law following the opposition's proposal—this was probably the only Bill which was a passed with the agreement of all the Parliamentary parties. The agreement was that the pitbulls be castrated and that henceforth such dogs must be kept secure—but really, and seriously, and not just on the word of the owner, but officially checked yearly, for which purpose the authorities can enter a place that is registered as a pitbull residence. Invariably, they cannot just walk into a place

that is not registered, because how would they know that what is roaring behind the door is the kind of dog that authorises them to enter? Veterinarian representatives (of the opposition) explained in vain that the solution was not strict enough, that the Bill drafted with the cooperation of the Ministry of the Interior responsible for the matter until now and the dog breeders' interest groups was unassailable, since the Bill came to the conference table as a joint opposition-government draft in the spirit of fervently desired national unity and collaboration.

(In the spirit of the rule of law, the legal basis for the official action and the government regulation was found in the Act on State- and Public-Security. This is a profoundly state-socialist decree of statutory force, regrettably from the vintage communist year of 1974 (1974. évi 17. tvr).) As the opposition member who introduced the Bill put it, if the state cannot protect its citizens from the canines, then "it values them less than a dog", and here "the state's reputation before its citizens is at stake".

Protection against dogs, or at least against pitbulls, was at last characterised as a task of the state. In the spirit of the rule of law the issue had to be classified as a matter of life protection, otherwise the Parliament may not interfere in dog-keeping, this being, at least according to the government, a question for local municipal administration.

So the new law applies nationally: it decrees that the pitbull and its hybrids are dangerous dogs in the entire country; in addition, the local government administrator *may* classify as dangerous a dog that has caused a person or an animal serious injury—that is, injury that takes more than eight days to heal and can only be remedied with plastic surgery. Breaking a bone likewise will do. (The criminal law uses criteria milder than this as a basis to ascertain serious bodily injury, which in certain circumstances is enough for imprisonment.) Administrative unpleasantries await the dog who is classified as dangerous: from now on he must wear an "I am dangerous" sign on his collar (which perhaps in a dog's eyes is not particularly defamatory). In the course of pronouncing the dog dangerous the *injured* person's conduct must be examined with "increased care", that is, if the injured person after his surgery still feels like becoming the object of an investigation. This time the presumption of the dog's innocence nicely prevails.

But in the case of a second attack there is no longer a presumption of innocence. The extermination of the animal must be ordered, provided that there is a valid authorisation. Which is a condition for tying up the animal as well. And if it disappears meanwhile? Legal security comes first. The legislator, however, recognised with great practical sense the danger of a wild dog—and prescribed an accelerated procedure. They can take the dangerous dog away as early as two weeks after the second biting. Assuming, that is, that the injury caused is once again serious, because if after the second attack the victim does not need plastic surgery, then the dog gets off with castration. (In the rule of law state punishment of the offender is "proportionate and progressive"—although I also stud-

ied this in Soviet criminal law, where in the textbook the unforgettable saying could appear that "the prison is the school of free social life". What I wouldn't give to be able to point out that the post-communist conceptions relating to the rule of law and legal order originate from the readers of those kind of textbooks and perhaps from their very authors!)

According to the news reports, in the year following the enactment of the law one pitbull was registered (and, understandably, castrated). By the way, next to the murderous pitbull whose act triggered action in Parliament there lived two more of them, but they moved with their owner to an unknown place. The promulgation of the law, it seems, eradicated the pitbull danger from Hungary, just like long ago the bloodhounds also disappeared if the town crier announced that Bloodhounds must be captured. From now on breeders will register pitbulls as a different breed. (But supposedly they really are disappearing, because the proprietors of the pitbull fights took the beasts to Transsylvania, where it is cheaper to keep them and there is less of the official fuss.)

The gods of the rule of law state—in case they are thirsty—are now appeased. (Human blood has flowed.) The rules were brought in defence of order, and on the basis of the requisite authorisations. The freedom loving state, out of tiredness and incompetence—despite its disposition and nature otherwise—made the oppressive tendency of the embarrassing predecessors (the 1974 state security statute, the 1968 metropolitan ordinance) disappear without a trace. It has no power to maintain its own order, let alone the order of voluntary organisations. In the end, in the name of rationality, we also succeed in conforming to the Rottweiler, with a local curfew. Order also comes into being when the state lets its citizens make their own arrangements for self-defence and self-preservation. Leaving us to ourselves is not neglect, but even if it is neglect, only the helpless person regrets it. Those who are well off arrange for their own protection. In fact, they even arrange the enforcement of their own rights. Mancur Olson demonstrated that systems without corruption and arbitrariness are historical anomalies and chance occurrences. Likewise, social order and peace are happenstances. Since it is not possible to force chance, it is therefore not necessary to do so. At least in post-communism.

According to the starting point of the Hungarian idea of the rule of law, the dog presumably is not dangerous, not bothersome, and in turn, belongs in residences and parks (at least in Budapest). It is the realist, at any rate, who makes order: in light of such poverty where would the public funds be to hinder the dogs? Who can afford a park guard? And even if hiring park guards would also alleviate unemployment, what could the guard do? And what would he say, how could he argue as a participant in the rational discourse, if the owner of a German Shepherd came? Why is it that from the point of view of the rule of law state and its order the question of principle is not, what are the authorities allowed to do, but instead, what can they say? What are the guard's arguments? What kind of legitimacy can his actions have? The owner, in case he doesn't speak the language of force, believes—and he is not alone—that he has rights

here in the park and on the sidewalk. Not more, however, than those without a dog. Everyone is free—at last—to keep a dog. To start a business, such as dog-breeding, is a constitutional right. Beyond the right of unlimited initiative the post-communist citizen does not have much respect for freedom, because for the most part it comes with annoyances and liberal blustering. He understands freedom primarily as "being left in peace"—or rather as leaving him in peace. This is what people did not like in the otherwise popular time of Communist Party *General Secretary* Jànos Kádár, that the state stuck its nose into everything. The spirit of freedom is lacking from this "leave in peace" sense of freedom. Spirit in other respects is also unfamiliar here. The people do not believe in anything, only in the tangible. Empiricism is not the realm of freedom.

There is a small problem with dog ownership: the master doesn't pay tax on his dog. Not even on the profit from breeding it. But what does this have to do with parks or with paying (with public funds) to clean up the dog mess? The evasion of taxes is simply part of the post-communist order, and for that matter, part of the accumulation of property. Otherwise ownership would be too expensive. There is no dog tax, however—democracy abolished it in the name of social justice and rationality. It would have been an ugly thing to collect tax from the pensioner who already suffers greatly and who is completely ruined by inflation. Public sentiment goes to the dogs; the revenue from the canine tax does not even cover the cost of collecting it. When the communists wanted to collect a dog tax, the only result was that the forests became full of abandoned dogs. The dog tax is not humane. People do not register their dogs anyway, because they begrudge the money even for the vaccines. If there were taxable consequences, they would not go to the veterinarian at all. Then there would be rabies, tuberculosis, distemper—all kinds of infections and epidemics.

So we cannot accept a dog tax, even for the sake of the dogs.

The dog owner believes that everything is due to him that is due other owners. He only wants equal rights. Why should he stay away, why should he pull on the leash, why should he keep his dog back, why should he cause distress and discomfort to his most faithful friend, who of course does not harm anyone, at least not good people? It would be a lesser problem for others to get out of the way. (The dignity of loafing around, the comfort without empathy, the Pepsi feeling, is not a Hungarian invention. But it discovered very "cool" soil here among us.)

Why should a dog avoid the children's sand box? Why should a dog be less valued and entitled to less than another person's child? ("You'll notice that he does not even bother my own children!"). Is the dog owner perhaps inferior to other pedestrians? Is he not entitled to the enjoyment of the publicly-funded park just like everyone else? Why should it not be the dog's right to have a nearby dog path, just like the children have a playground? They should make an appropriate dog path nearby and every problem would be solved. (The regulation in Budapest makes the cleaning of the dog paths an obligation of the Public Works department—without extra compensation. All they need now is to col-

lect an entrance fee on the dog path! If there is a legal ground for it. And if they dare.) That the Constitution speaks of the state's obligation for the protection of children, but is silent about dogs, is a viewpoint which we definitely cannot initiate into the discourse of the park warden and the dog owner pulling his hundred pounds' pet out from the basement.

Just like you or someone else can go around in your car, get up on the sidewalk, and defecate, so can a dog go around, get up on the sidewalk, and defecate, right? Both are property, right? (And then the question of humanism does not come up.)

In the past, cars illegally pushed their front ends up onto the sidewalk, primarily in response to the fact that they multiplied along with the impoverishment following the collapse of communism. They did not have many choices; in the absence of parking places they had to stop somewhere. The situation is no better in Italy. The authorities allowed it; after all, common practices cannot be treated as an offence. Even if the driver is simply lazy or in a hurry and doesn't look for a parking space, or begrudges the money for it. As with dog-keeping, there are no longer enough private resources or will for the additional costs of driving (the road, environmental protection, brakes, insurance)—and this determines the limits of the disorder. Baby carriages and disabled persons solve their problem as they wish; besides, there are less and less children while our car pool happily multiplies.

But if here and now the car proves the superiority of the physical facts incidental to ownership, if the order dictated by automobiles (the truth dictated by necessity) sends a bad message about the possibility of rearranging order, the weakness of the authorities, and the unity of the accomplices, all of this does not necessarily have to materialise in the matter of dog-keeping. The automobile is, after all, a more "massive" phenomenon; it is more "a part of" the sustenance of the middle and lower-middle classes, the transport of the standard life. Anyway, the car owner has less choices; in certain cases he truly needs a car for his work and therefore he truly needs to park. But why does a person need to have a dog? Because "Rome must sail"? But if someone already has one, why can't he pick up its mess? Moreover, like second-hand cheap car alarms disturb the quiet of the night (indeed, there are bark-free hours) and are not fined for it, dogs who are cheaply kept, and for that reason (also) are neurotic, can howl with impunity. Just like smashing a jammed alarm is an unlawful trespass (in Italy as well), people would probably be frightened by the spectre of vigilantism if a person walking down the street seized his air gun. He who protests is an idiot. Besides, the complaint about disorder and the physical deterioration are just fastidiousness and cultural arrogance; disorder became question of viewpoint. Conceptions of order compete; the order of simplicity prevails. Is it only possible to get between the wall and the car parked on the sidewalk by walking sideways? Well, what of it? There are worse things. Besides, *walking* around the filthy city is such an out-dated idea. One drives down in the morning from the exclusive hills of Budapest, and when finished, one rushes back.

In the *Litte Prince* by Saint-Exupery, the fox says to the prince: "because you tamed me, you bear the responsibility for me". The *Little Prince* is the favourite

(perhaps only known) book of the tiny Hungarian middle class.

The dog owner is no prince. He does not want to take on—responsibility, not so much in relation to his dog, but for his dog, in relation to everyone else.

Our judges have learned, already under communism, and even since then from the textbooks bequeathed to us from that time, that the law cannot be indifferent to society. Even the hunters, who otherwise enjoy all sorts of privileges (exemptions), may not just shoot down a wolf-looking stray German Shepherd. On the other hand, the hunters do not have to exterminate the "wandering bloodhounds". A court of the second instance ruled that a Hunters' Society is not only qualified to shoot a stray dog, but is required to do so because of the dangerous situation that has developed: "they have the right to the most effective protection, to the use of weapons, as long as the plaintiff [on whose farm the stray dogs destroyed the calves] so to speak surrendered to the bloodhounds". The Supreme Court, however, already in 1990 observed that a dog is not a wild animal, therefore in relation to it—in contrast to that which concerns the damage caused by a wild animal—the Hunters' Society does not have an obligation. (According to a 1971 regulation, hunters' societies, which have an exclusive right to hunt in their respective territories, are only required to exterminate stray dogs in the interest of management of the wild. The Supreme Court ruled that failure to do so which results in damage to third parties does not result in liability. The Hunters' Society has the exclusive right to shoot in its area; however the country people themselves should solve the problem of protecting their calves).

The dog owner is not interested in what kind of damage his dog causes, it is only to the dog that he has obligations: he has to ensure that the dog feels good. Humanism would wish that others would also tolerate this. What is more, in view of the post-communist citizen it would be up to the state to contribute to the general well-being of the dogs. A dog needs exercise! And where should the owner exercise the dog, if the so-and-so's didn't grant an appropriate dog path? Perhaps on the asphalt, on the sidewalk? That would be dangerous, because the cars might get close to them. Besides, the park grass is more pleasant and more natural.

They told the poor children: "you have a right to it!" (Pascal). And they believed it and they thought that therefore there are no limits. But because they were children, they were powerless and mostly poor. Today a lack of limits implies limitless leisure. And if this lack of limits is accompanied by diligence, it is just that much worse. They do not understand that the absence of limits also has a boundary—that of not causing harm to other people. Simple and stupid—they do not realise the effects they have and what it is that affects them. The simple-minded Raskolnikov of today, the man without consequences, beats an old woman over the head and wonders (if he ever wonders at all): how on earth did she kick the bucket just from me smacking her on the bald crown of her head?

And all of this trickles through into the law; after all, there is democracy; besides, "legal regulation has to conform to life", and furthermore, "it is impossible to create law in opposition to the people's will". We could enumerate the distorted, poorly-applied maxims and sloppy platitudes from morning to night. But

it is a fact that in relation to the dog owner the other apartment owners can only be a humble petitioner in Budapest municipal law. There is no presumption of risk or nuisance associated with keeping a dog; in fact, verification of a neighbour's complaint is not enough—the local government administrator has to find a public interest more compelling than nuisance in order to restrict or remove the dog. The fact that the neighbours' private property or proprietary right is damaged is not enough. Ownership is considered freedom of action, and not a quiet, passive pleasure. The property belongs to him who dares to use it. Property cannot count on protection, however. The property right of entrepreneurs and conquistadors is this: they were first, the recognition goes to the first impertinence. So it follows (and this is troubling for the social peace) that whoever helps himself can hope to be undisturbed in his effort. Let's use spray, poison, killer dogs. This understanding of freedom of property leads to the law and order of the stronger dog. If the stronger dog dictates the fundamental order, then every formality of the rule of law restrictions associated with it only brings about this—the law and order of the stronger dog. Speaking more bluntly, the authorities can do me a favour. But of course in the absence of resources the authorities will not kiss up, nor clean up, and so then the shit remains, until the ever less frequent rain.

The Hungarian constitutional revolution somehow did not discover that property can have associated duties, and most importantly, that the natural limit of the freedom of property is the freedom of others.[1] The mention of obligation raises suspicion: in communism social obligations were followed by general oppression; after communism the utterance of these words is followed by urgent filling of someone's pocket.

The death of state socialism slowly liberated the people, but the remaining over-grown state, precisely as a result of its size, is ineffective and slothful, and therefore is not capable of doing anything with the everyday reality of libertinism. The people expect a great deal from the state, and the state from itself, but the state, being neither Croesus nor Hercules, inevitably turns out a weak performance. It cannot act, but it does not want anyone else to either.

The market and private interest need to hurry for help. A happy end!—enclosing our park (and several others as well) using huge amounts of public funds. After all, we live in the age of big business, and even in post-communism the municipality business needs form a particularly large amount of public procurement. The *fence* is a sign of serious proprietary thinking. Supposedly, there will also be a guard. Otherwise the dog owners will not tolerate being forced into the ghetto of the designated dog-running area.

Tibor Déry's visionary book, *Mr. G.A. in X*, depicts the flip side of a line by the communist poet Attila József. "Freedom, give birth to order for me!" prayed the poet. In the city of X however, complete freedom—limited only by how much one can exert himself—creates chaos. (These two Bolsheviks foresaw something.) But the citizens of X at least dispose of civilian virtues, that is,

[1] Although the Hungarian constitution is full of accidents, at least it is paradigmatical, if also accidental, that in relation to property there is nothing said about social obligations.

they are eternally patient. Post-communist freedom dropped into the laps of disorderly people or people who had been made disorderly (while they sat around waiting for their ship to come in). They did not liberate themselves, they just escaped. And this escape frames the state, the conventional, moral, and economic institutions. The dogs of neurotic masters are also neurotic.

As I read in the trade journal *Dog*, the dog is an "unnatural spasm, an ethological mistake". According to many people, the Soviet reign made its subordinates such spasms; according to others, wolf-capitalism (the mother of dog-capitalism) makes our compatriots inhumane. No matter how it is, this spastic man is his own (and his dog's) master. This anthropological potentiality defines the institutions which the spasms establish. So why would these institutions, built by these people, be better able to control them? Because we desire order (besides however much we may also love this domination). The coerced, planned order was unpleasant. According to the market idolaters, order supposedly comes into being from free, unplanned, uncoerced deeds. Perhaps this is so, if the free deeds can conform to some rational system of ordering.

Was it this kind of freedom, dog freedom, that gave birth to disorder? Phenomenologically, it is undoubtedly on the same level as the cracked asphalt, covered with dog excrement.

According to Professor Gáspár Miklós Tamás, a leading observer of post-communism, it is the narrow-minded petty bourgeois who is returning the compliment when his response to freedom in the new era is that the street is dirty. However, no pettier a bourgeois than Nietzsche believed that the cleanliness of the streets in Torino is worthy of adoration.

The streets of Budapest, however, really are dirty. Italian fascism, which was much more oppressive than the late Kádár era, functioned as a petty bourgeois dictatorship similar to Kádár's. While it existed, but afterward as well, everyone recognised, in a truly petty bourgeois way, that "at least the trains run on time". Since Italy has had the freedom to strike, labour unions, and job security, and mostly because of these things, the trains are late. Or they do not even depart. The Hungarian trains, on the other hand, were late during the dictatorship of the petty bourgeois and are late during democracy. Blame the rails.

For someone who is pro-freedom, it is especially embarrassing to cry for order. What is more embarrassing, whoever pants after order usually ends in yelling for the police. Is the dog excrement getting to me?

The dogs, without muzzles or leashes, produce filth and fear, with the help of the intentional impotence of those authorised to maintain law and order. The masters and the "victims" (both with and without dogs), live in fear like stray canines.

The stray canine, writes *Dog* magazine, is a double outcast. Over a period of six thousand years people have removed the dog from his natural environment and his pack, and a few weeks after birth they separate the puppy from his mother and siblings. After this the abandoned dog is removed from his remaining substitutive human company as well.

What kind of order can stray dogs create?

12

Globalization and the Fate of Law

WILLIAM E. SCHEUERMAN*

Few notions within modern political and legal theory have been more widely accepted than that capitalism and a legal order based on clear general norms, in conjunction with relatively formalistic modes of legal decision making, necessarily go hand-in-hand. From John Locke to Max Weber, the existence of an "elective affinity" between capitalism and the rule of law represented a core element of liberal theory. Friedrich Hayek's recent attempt to weld a formalistic model of the rule of law directly onto a defence of free market capitalism is merely an exaggerated statement of certain themes already found within classical liberalism.[1]

Liberalism's opponents have also presupposed the existence of an intimate kinship between a market economy based on private property and legal institutions promising substantial legal calculability. No less so than Locke or Weber, Marx believed that the formalities of "bourgeois" law were closely allied to modern capitalism, while "post-Marxist" radical jurists (for example, proponents of Critical Legal Studies within North America) now often similarly reproduce the traditional view of a special relationship between capitalism and the rule of law.[2]

The main difference between the liberals and their critics is that the latter typically rely on the notion of an elective affinity between capitalism and the rule of law to discredit both institutions, whereas the former still see market capitalism as indispensable to even minimal guarantees of legal certainty. In this spirit, liberal politicians and publicists today characteristically see market reforms (for example, in China or Eastern Europe) and the rule of law as two sides of the same coin: capitalism and the rule of law allegedly require each other, and thus market reforms must be accompanied by legal changes pointing in the direction of the liberal rule of law.

In light of this surprising consensus, as well as an impressive body of historical scholarship documenting the intimate links between economic and legal liberalism in modern history, it might seem odd to try challenge the idea of a

* I am grateful to David Dyzenhaus for providing incisive critical comments on an earlier version of this essay, as well as to Steven Young, who has helped me think through the issues raised in this essay.

[1] Friedrich A. Hayek, *The Road to Serfdom* (Chicago: University of Chicago Press, 1944).

[2] In this vein: Roberto Mungabeira Unger, *Law in Modern Society* (New York: Free Press, 1976).

kinship between capitalism and the rule of law.[3] Yet here I hope to do just that. In my view, the traditional belief in an elective affinity between economic liberalism and the rule of law obscures the manner in which the ongoing emergence of a global capitalist economy *threatens* core features of the rule of law. Contemporary global capitalism is distinct in many ways from its historical predecessors: economies driven by huge multinational corporations, rapid-fire electronic communication and high-speed transportation, and the emergence of supranational economic blocs, represent a novel development in the history of modern capitalism.[4] The relationship of capitalism to the rule of law is thereby transformed as well. A hitherto unrecognised dialectic has been at work in modern capitalism: by incessantly revolutionising the time and space horizons of economic action, capitalism tends to diminish its reliance on a relatively robust model of the rule of law. The capitalist-induced "compression of space and time" systematically limits capitalism's dependence on consistent and general forms of legal decision-making. The legal infrastructure of contemporary economic globalisation suggests that this historical dialectic is now coming to fruition, as traditional modes of liberal law decreasingly figure in the operations of the global economy.

I begin by recalling the conceptual outlines of the traditional notion of an elective affinity between economic liberalism and the rule of law (section I). Then I suggest why capitalism's revolutionary implications for the time and space horizons of economic action today render important features of the traditional story anachronistic (section II). Finally, I draw some tentative political lessons (section III). My aim here is *not* to salvage a bankrupt intellectual tradition in which legal institutions were analysed chiefly in terms of their social and economic functions; that tradition has rightly been discredited. Yet I do believe that to examine legal practices "in monadic isolation from their social and economic context is —for many purposes— like playing *Hamlet* without the Prince".[5]

[3] On the American case: Morton J. Horwitz, *The Transformation of American Law, 1780–1860* (Cambridge: Cambridge University Press, 1977); on Britain: P.S. Atiyah, *The Rise and Fall of Freedom of Contract* (Oxford: Clarendon Press, 1979); on Germany: Franz Wieacker, *Industriegesellschaft und Privatrechtsordnung* (Kronberg: Scriptor, 1974).

[4] For a useful discussion of the concept of economic "globalization": Paul Hirst and Graham Thompson, *Globalization in Question* (Cambridge: Polity Press, 1996). My discussion here focuses chiefly on the legal implications of three of its features: (1) the growing importance of multinational corporations (MNCs) within the world economy; (2) the transnationalisation of capital and financial markets, and (3) "compression of space and time". I neglect for now the pivotal question of how the growth of regional economic blocs (NAFTA, ASEAN, the European Union) impacts on legal development.

[5] Mirjan R. Damaska, *The Faces of Justice and State Authority: A Comparative Approach to the Legal Process* (New Haven: Yale University Press, 1986), p. 7.

1. AN ELECTIVE AFFINITY BETWEEN ECONOMIC LIBERALISM AND THE RULE OF LAW?

For my purposes here, it suffices to define the rule of law as requiring that state action rests on legal norms that are (1) general in character, (2) relatively clear, (3) public, (4) prospective, and (5) stable. According to the mainstream of modern liberal theory, only legal norms of this type assure a minimum of certainty and determinacy within legal decision-making, contribute to achieving equality before the law, guarantee the accountability of power-holders, and promote fair notice. The rule of law renders the activities of power-holders predictable and thereby makes an indispensable contribution towards individual freedom.[6] Elsewhere I have tried to argue that a principled commitment to a traditional model of the liberal rule of law of this type need not unduly hinder the quest for a generous welfare state. Despite the existence of an impressive intellectual consensus asserting the contrary, greater social and economic equality and more formal legality *can* and *should* go hand-in-hand. In my view, the democratic left would do well to support a relatively robust model of the rule of law, in part because scholars too often have obscured the potential perils to the socially and economically vulnerable of highly discretionary modes of judicial and administrative decision-making. Anti-formalism and social democracy make poor bedfellows.[7] By underlining the fundamental tensions between traditional forms of liberal law and the aggressive brand of economic liberalism now dominant within international economic affairs, the present essay represents a further step in my attempt to undergird an identifiably *social democratic defence of the rule of law*, in which the achievements of classical liberal jurisprudence are taken seriously.

Although often ignored by its critics, a commitment to the ideal of the rule of law hardly requires fidelity to a crude *hyperformalism* according to which the

[6] I elaborate on these familiar virtues of the rule of law during the course of my exposition below.

[7] William E. Scheuerman, "The Rule of Law and the Welfare State: Towards a New Synthesis" (1994) 22 *Politics & Society* 195–213. Also, my *Between the Norm and the Exception: The Frankfurt School and the Rule of Law* (Cambridge, Mass.: MIT Press, 1994), where I defend a social democratic model of the rule of law by relying on the insights of the early Frankfurt School jurists Franz Neumann and Otto Kirchheimer. In a powerful essay (included in this volume), Henry Richardson criticises my position in part by offering a cautiously positive gloss on forms of post-traditional, relatively decentralised modes of judicial and administrative making. For now, let me just comment that I am not in principle opposed to experiments like those endorsed by Richardson; the social democratic model of the rule of law that I have tried to defend does not preclude them. Yet I am far more sceptical than Richardson concerning their potential advantages as an instrument for warding off the undue influence of well-organised forms of concentrated private power. Particularly in the era of globalisation, in which large corporations exert an enormous influence on political and legal decision-making, the problem of taming economic power must take a prominent place in legal theory, and I continue to believe that relatively traditional forms of general law have an important role to play in that quest. My worry is that Richardson (1) still tends to downplay some of the impressive virtues of relatively traditional forms of law and (2) places too much faith in alternative legal forms just as likely to exacerbate the side-effects of economic and social inequality as reduce them.

rule of law allegedly implies that there is only *one determinate* answer to every legal question, and every manifestation of judicial or administrative autonomy represents an attack on the principle of legality.[8] Even the most cogent legal rule can be interpreted in relatively distinct ways, and it sometimes makes sense for a polity to delegate discretionary authority to courts or bureaucrats. Yet even if legal materials often fail to determine a single correct answer, clear and cogent legal norms can provide a framework in which a relatively limited set of acceptable answers is delineated. A measure of indeterminacy within the law is unavoidable, but indeterminacy can be contained and managed by legal norms possessing the attributes of generality, clarity, publicity, prospectiveness, and stability. In short, the rule of law is consistent with what we might describe as the *limited indeterminacy thesis*, according to which defenders of the rule of law need not endorse exaggerated conceptions of legal certainty and regularity.[9]

Nor does a defence of the rule of law in the sense described here require subscribing to the basic tenets of legal positivism, at least if positivism is seen as necessarily entailing a strict delineation of legality from morality. As Judith N. Shklar noted many years ago, "[i]t is. . .one thing to favor the ideal of a *Rechtsstaat* above all ideological and religious pressures, and quite another to insist upon the conceptual necessity of treating law and morals as totally distinct entities".[10]

Within the history of modern legal and political thought, many authors committed to a relatively formalistic model of the rule of law, in which judicial and administrative discretion were supposed to be kept to a minimum, refused to endorse an airtight separation of law from morals. For my purposes in this essay, the continuing intellectual battle between positivists and anti-positivists is of secondary significance, given the fact that both positivists and anti-positivists, though for different reasons, can endorse a model of law as ideally

[8] As far as the history of modern legal thought is concerned, some (especially early liberal) theorists did come close to defending an overstated model of legal formalism, whereas many (more recent) theorists did not.

[9] I develop the notion of "limited indeterminacy" in the introduction to my *Carl Schmitt: The End of Law* (Lanham, MD: Rowman & Littlefield, 1999). In doing so, I borrowed from Lawrence Solum's enlightening essays on different forms of indeterminacy within the law: "On the Indeterminacy Crisis: Critiquing Critical Dogma" (1987) 54 *University of Chicago Law Review* 462–503 and "Indeterminacy," in Dennis Patterson (ed.), *A Companion to Philosophy of Law and Legal Theory* (Oxford: Blackwell Publishers, 1996), pp. 488–502.

[10] Judith N. Shklar, *Legalism: Law, Morals, and Political Trials* (Cambridge: Harvard University Press, 1986), p. 43. At least implicitly, Shklar was responding to writers like Ronald Dworkin, for whom the failure of the positivist quest to guarantee a strict separation of morality from legality means that a traditional "rule-centered" model of law is best replaced by a rights-based interpretation of the rule of law (Dworkin, *Taking Rights Seriously* (Cambridge: Harvard University Press, 1977); and *A Matter of Principle* (Cambridge: Harvard Univerity Press, 1985)). For Shklar's criticisms of Dworkin, see "Political Theory and the Rule of Law", in her *Political Thought and Political Thinkers* (Chicago: University of Chicago Press, 1998), pp. 32–6. I think that Shklar was right to be sceptical of this move within Dworkin's theory. It is, of course, possible to defend the rule of law, along with relatively formalistic models of judicial and administrative action, without ignoring their "moral" or "ethical" qualities.

possessing the attributes described above.[11] Both positivists and anti-positivists will also be able to identify sufficient reasons to worry about the pervasive anti-formal trends within global economic law that I describe later in this essay.

Space restraints prevent me from offering a detailed survey of the myriad ways in which modern political and legal thought conceived of a special relationship between a capitalist economy and the rule of law. Yet it does make sense to try to recall three core arguments underlying the notion of an elective affinity between capitalism and the rule of law.

The calculative ethos of the modern capitalist enterprise arguably constitutes one source of the kinship between capitalism and the rule of law. In a line of argumentation developed most completely by Max Weber, "exact calculation" is conceived as reigning supreme within modern capitalism. Characterised most basically by a "systematic utilisation of goods or personal service" in which "calculation underlies every single action of the partners", modern capitalism depends on highly developed forms of accounting and book-keeping, the separation of business from household activities, and formally free labour, each of which makes a vital contribution towards achieving predictable forms of economic activity promising maximum control over the natural world.[12] For Weber, the modern entrepreneur is a sober, bourgeois character, embodying a demanding "asceticism [that] was carried out of monastic cells into everyday life".[13] The discipline of the entrepreneur corresponds to the imperatives of an increasingly calculable and predictable economic universe in which "the technical and economic conditions of machine production . . . determine the lives of all the individuals who are born into this mechanism . . . with irresistible force".[14] In Weber's account, in the pre-modern "capitalism[s] of promoters, speculators, concession hunters. . .above all, the capitalism especially concerned with exploiting wars", entrepreneurial activity was often "irrational" and adventurous, as capitalists pursued profit by reckless speculation, piracy, or even force. In earlier forms of capitalism (for Weber, an economic system having many distinct historical variants) the principle of "exact calculation" was anything but supreme.[15] But in modern "rational" capitalism, the entrepreneur allegedly trades in his more romantic traits for the self-possession of the cautious, calculating businessman. Just as modern capitalism relies on the principle of exact calculation, so too does its leading feagure, the modern entrepreneur, come to embody a disciplined, systematic and calculative ethos.

[11] Compare, for example, Joseph Raz, *The Authority of Law* (Oxford: Clarendon Press, 1979), pp. 210–29, and Lon Fuller, *The Morality of Law* (New Haven: Yale University Press, 1969), pp. 33–94. Of course, this is not to deny the obvious point that fundamental jurisprudential differences generate differences in competing models of the rule of law; Raz and Fuller obviously disagree in many ways. But for my purposes here, such differences are of relatively peripheral importance.

[12] Max Weber, *The Protestant Ethic and the Spirit of Capitalism*, (Talcott Parsons (trans.), New York: Routledge, 1992), pp. 18–19.

[13] Ibid., p. 181.

[14] Ibid.

[15] Ibid., pp. 20–1.

In light of this picture of the "ascetic" modern capitalist entrepreneur, it is easy to see why Weber, and so many influenced by him, believed that modern capitalism requires the rule of law.[16] Just as modern capitalism makes "exact calculation" supreme, only modern "rational legality", defined as resting on general, clear, and well-defined concepts and norms, provides for optimal legal calculability. Modern capitalism aspires to achieve maximum predictability in economic affairs; a legal order devoted to assuring maximum calculability represents its natural institutional complement. In Weber's famous account of modern law, legal decision-makers ideally were to approximate "an automaton into which legal documents and fees are stuffed at the top in order that [they] may spill forth the verdict at the bottom along with the reasons, read mechanically from codified paragraphs".[17] The "machine-like" character of modern rational legality corresponds directly to the predictability of modern capitalist "machine production". Although obviously exaggerated, Weber's account of modern law's "mechanical" features captures the gist of the modern liberal view that the rule of law counters unnecessary unpredictability. Generality within law protects against irregularity by demanding that like cases are treated in a like manner. Clarity serves the same function, for vague and incoherent laws often are applied and enforced in inconsistent ways. Stability similarly helps achieve calculability as well; rapid or confusing changes within the law contribute to unpredictability and uncertainty within its application and enforcement.

Virtually all forms of organised economic activity exhibit some minimum of calculability and regularity, since it is hard to imagine how regular, ongoing economic production or exchange would be possible without some element of orderliness and regularity. Even a pirate ship probably rests on a normative order and a measure of rule-like behaviour. But special to modern capitalism is that it makes the quest for predictability and calculability all-important:

> "It is not enough for the capitalist to have a general idea that someone else will more likely than not deliver more or less the performance agreed upon on or about the time stipulated. He must know exactly what and when, and he must be highly certain that the precise performance will be forthcoming. He wants to be able to predict with certainty that the other units will perform".[18]

For Weber, the capitalist preference for "precise performance" in economic relations leads him to prefer an equally precise legal environment maximising the chances that his expectations will be satisfied. And his desire to "predict with

[16] Weber's formulation here has exercised an enormous impact on continental legal theory. The list of writers influenced by him on this point would read like a Who's Who? of twentieth century jurisprudence.

[17] Max Weber, *Economy and Society* (Guenther Roth and Claus Wittich (eds.), Berkeley: University of California Press, 1979), vol. I, p. 657.

[18] David M. Trubek, "Max Weber on Law and the Rise of Capitalism" (1972) 3 *Wisconsin Law Review* 743.

certainty" makes him an ally of legal forms (for example, clear, calculable, and enforceable contracts) tending to reduce economic uncertainty.

Although often ignored, it is striking that this view relies on a specific —and probably controversial— model of the capitalist enterprise and its key figure, the entrepreneur.[19] If the capitalist entrepreneur could rest satisfied with a rough or approximate sense that his expectations were to be fulfilled, legal forms procuring something less than the machine-like certainty promised, in Weber's view, by modern rational legality, might suffice for him. Perhaps a somewhat less airtight, predictable system of law than that offered by a formalistic model of the rule of law could serve the capitalist well enough. Maybe capitalism and traditional legal forms exhibiting a limited degree of systematisation and formal rationality would be able to coexist successfully.[20] Of course, Weber himself probably rejected this possibility, in part because of the weight he placed on the role of growing predictability and calculability within his broader vision of the "rationalisation" of Western modernity.[21] Even his account of the common law hence tends to underline the manner in which it embodies "rational" elements structurally similar to those found within the formalistic legal codes of the European continent. Like their rivals in France and Germany, common law systems allegedly underwent a systematisation that functioned to assure the legal calculability necessary for modern capitalism.[22]

A second defence of the notion of an elective affinity emphasises the protective functions of the rule of law for the modern entrepreneur. Long engaged in a fierce battle with the legacy of European Absolutism, early liberal theorists (including Locke, Montesquieu, Beccaria, Voltaire, and Kant) conceptualised the rule of law as a puissant weapon against political despotism and economic paternalism. Legally unregulated, arbitrary government was typically pictured as posing a threat to every manifestation of individual freedom. Nonetheless, early liberalism's tendency to conceive of the individual chiefly as a proprietor often led it to place a special emphasis on the dangers of political and economic

[19] I develop this point in section II.

[20] This is one of the arguments made by Harold J. Berman's provocative *Law and Revolution: The Formation of the Western Legal Tradition* (Cambridge: Harvard University Press, 1983), which criticises Weber for exaggerating the amount of legal formalism requisite to the emergence and flourishing of modern capitalism. At least to the extent that we both ultimately question the existence of an elective affinity along the lines described by Weber and many other modern writers, Berman's study complements my own here. As my comments at the beginning of this section should suggest, however, I disagree with Berman's endorsement of relatively traditional, non-formal modes of law.

[21] For Weber, essential to the "rationalization" of modern Western society is that it drives "magic" from the world, while systematically rendering social and natural processes increasingly predictable and calculable in character. Modern capitalism and the rule of law are simply two interrelated components of this process. To the extent that predictability is built in to Weber's concept of rationalisation, it is unsurprising that he tends to emphasise (and probably exaggerate) the growing predictability of both its economic and legal manifestations.

[22] For more recent defences of this line of interpretation: Franz L. Neumann, *The Rule of Law: Political Theory and the Legal System of Modern Society* (Leamington Spa: Berg Publishers, 1986), pp. 239–52; Otto Kahn-Freund, "Einführung", in Karl Renner, *Die Rechtsinstitute des Privatrechts und ihre soziale Funktionen* (Stuttgart: Gustav Fischer Verlag, 1965), pp. 8–16.

despotism to commercial life. In this view, political arbitrariness is simply incongruent with the successful operation of a modern commercial economy. Inconsistent and irregular state activity make it difficult for proprietors, for example, to engage in necessary forms of long-term private planning, in which economic expectations have a reasonable chance of gaining satisfaction. Why invest when the spectre of unforeseen state activity risks wiping away any economic advantages to be gained by doing so? Similarly, when property rights or contracts are unsettled as a result of an unreliable state administration, even simple economic transactions become unsettled and unduly problematic as well.

It is hard to deny the underlying strength of this early liberal insight. In the aftermath of the emergence of the modern state and its awesome monopoly on organised violence, no institution has posed a greater threat to the quest for economic certainty than the state's bureaucratic apparatus: since Machiavelli, political theory has been preoccupied with the task of showing how both the sovereignty of the modern state and political and economic freedom can exist together. From this perspective, we can quickly grasp why the rule of law so often represented an unambiguous good for those committed to the emergence of modern capitalism. *Pace* laissez-faire ideology, liberalism was not in principle opposed to state action within the economy per se; throughout the history of capitalism, the state has played a substantial role in economic affairs. But liberalism understandably was opposed to forms of state action likely to generate unnecessary economic uncertainty and unpredictability. Clarity and publicity within law assure that entrepreneurs gain fair notice of when and how governmental officials are to intervene. Secret or retroactive legal norms make it difficult for entrepreneurs to know how state agents are likely to act, and thus a liberal legal order best steers clear from them. Generality and stability similarly contribute to the accountability of power-holders by helping make sure that they at least act in a consistent way. When Hayek famously wrote in *The Road to Serfdom* that "stripped of all technicalities this [the rule of law] means that government in all its actions is bound by rules fixed and announced beforehand —rules which make it possible to foresee with fair certainty how the authority will use its coercive powers in given circumstances, and to plan one's individual affairs on the basis of this knowledge", it was primarily this protective function of the rule of law that he probably had in mind.[23]

Finally, the plausibility of the idea of an elective affinity between capitalism and the rule of law can be traced to the time and space horizons of economic action within modern capitalism. Anthony Giddens is right when he insists that social and political theory needs to give "conceptual attention to the timing and spacing of human activities", and that the mainstream of modern thought generally has failed to do so.[24] Indeed, explicit references to the problems posed by

[23] Hayek, *The Road to Serfdom*, n. 1 above, p. 72.

[24] Anthony Giddens, *The Nation-State and Violence* (Stanford: Stanford University Press, 1987), p. 12. Within social theory, attention to this issue has grown dramatically in recent years. For a

the temporal and spatial context of economic action are rarely found in modern political and legal thought. Yet implicit assumptions about the space and time horizons of economic activity play a crucial role in traditional thinking about the rule of law.

Let us recall the seemingly trivial fact that economic action in the history of modern capitalism often entailed time-consuming and relatively risky forms of exchange. Distance potentially generates real uncertainty; think of the difficulties posed by the long-distance transport of goods before the advent of the steamship, railroad, automobile, or airplane. For the moment, imagine a merchant trading in the backwoods of North America in the late eighteenth century, whose business relied on long and risky voyages from a port city on the coast (Boston or Charleston perhaps) to the frontier (Kentucky or Ohio). From the perspective of our early capitalist merchant, one way to reduce economic uncertainty would be to make sure that laws impacting on his business remained unaltered by the time he completed his trek and returned to his home port from the rural hinterlands. It would make economic sense to seek tax laws unlikely to change during the course of his travels, a stable system of contracts, and many other relatively predictable legal norms and practices. Quick or sudden changes in the tax code —for example, new taxes on products traded by him— would be undesirable to the extent that they unnecessarily heightened economic uncertainty and made it difficult for him to plan his actions in a rational manner. Clarity and transparency in law would promote his understanding of the code and reduce economic insecurity. Its generality would work to prevent him from being discriminated against in relation to similarly situated merchants and traders. In short, familiar features of the rule of law would serve as a powerful tool for counteracting uncertainties generated by the distance and duration of economic exchange. A liberal legal system would not only faciliate capitalist exchange by securing private property and a system of free contracts, but also by dramatically reducing insecurities deriving from the time and space horizons of the merchant's economic environment.

Perhaps the point can be better illustrated by drawing a parallel to a widely-documented shift in the history of map-making and time-measurement. In premodern Europe, maps and clocks were clumsy and imprecise devices, less concerned with a rational calculation of time and space than serving didactic moral and religious purposes. Only in the Reformation and Enlightenment do we see the proliferation of maps and clocks based on a disenchanted view of nature as an object of conscious human manipulation and control. Revealingly, this quintessentially modern attempt to improve the rational management of the physical environment was inextricably linked to ever more precise, systematic modes of time and space measurement: increasingly exact and reliable systems of measurement based on quantifiable, homogeneous, linear units of time (culminating in modern mechanical watches and clocks), as well as maps in which

survey: John Urry, "The Sociology of Time and Space", in Bryan S. Turner (ed.), *The Blackwell Companion to Social Theory* (Oxford: Blackwell Publishers, 1996), pp. 369–95.

space was conceived in terms of abstract, uniform grids, lacking the elements of fantasy and religion that once had been paramount in the minds of medieval map-makers.[25] This seemingly mundane innovation nonetheless contained revolutionary implications: the dramatic shift in the time and space horizons of European civilisation, given concrete form within the history of modern cartography and time measurement, played a crucial role in the European conquest of the non-European world.

The development of modern law exhibits remarkably similar structural characteristics. Within modern Europe, the legal system was more and more conceived in terms of a set of abstract, formal, and general propositions, making up a systematic code to an ever greater extent free of traditional moral and religious overtones. Just as modern clock and map-makers carved time and space into homogeneous units possessing an ever more precise and systematic character, so too did modern legal reformers imagine a complementary legal universe consisting of rationally-ordered, uniform, abstract concepts and norms, as well as formalistic modes of decision-making ideally no less predictable than the operations of a modern clock or reliable than a good map. Enlightenment intellectuals often brought these related strands within modern thought together. For example, Voltaire was not only fascinated by the possibilities for rational urban planning inspired in part by a conception of space as consisting of abstract, uniform units, but famously fought to modernise law by basing it on transparent, abstract, universal propositions.

In light of this striking parallel, would it be too far-fetched to suggest that not only modern cartographers and clockmakers, but modern jurists as well, hoped to improve humanity's instrumental mastery of time and space by means of their contributions to modern culture? For now, my point is a modest one. Like rational maps and clocks, the modern rule of law implicitly rested on the aspiration to render both time and space rationally manageable. One crucial way in which it achieved this task was by reducing economic uncertainty based on the distance and duration of commercial exchange.

II. MODERN IMPLICATIONS FOR THE TIME AND SPACE HORIZONS OF ECONOMIC ACTION

If the idea of an elective affinity between capitalism and the rule of law requires attention to the time and space horizons of economic action, it becomes incumbent on us to acknowledge the ways in which modern capitalism incessantly transforms those horizons. The social theorist David Harvey is right when he describes capitalism as:

[25] David Harvey, *The Condition of Postmodernity: An Enquiry into the Origins of Cultural Change* (Oxford: Blackwell Publishers, 1989), pp. 240–59. On time measurement and the rise of modern capitalism: E.P. Thompson, "Time, Work-Discipline, and Industrial Capitalism" (1967) 38 *Past and Present* 56–97.

> "a revolutionary mode of production, always restlessly searching out new organisational forms, new technologies, new lifestyles, new modalities of production and exploitation and, therefore, new objective social definitions of space and time . . . The turnpikes and canals, the railways, steamships and telegraph, the radio and automobile, containerisation, jet cargo transportation, television and telecommunications, have altered space and time relations . . . The capacity to measure and divide time has been revolutionised, first through the production and diffusion of increasingly accurate time pieces and subsequently through close attention to the speed and coordinating mechanisms of production (automation, robotisation) and the speed of movement of goods, people, information, messages, and the like".[26]

Driven incessantly to search out new sources of profit, capitalism constantly overcomes geographical barriers and transforms technology, thereby revolutionising the time and space horizons of social life: "the history of capitalism has been characterised by speed-up in the pace of life, while so overcoming spatial barriers that the world seems to collapse inward upon us".[27] The reduction of turnover time (in production and marketing) is a decisive means for capitalists to increase profits. Especially during moments of intense competition and crisis, capitalists with faster turnover time are likely to outpace their competititors. The same can be said about capitalism's drive to overcome spatial barriers; geographical expansion into new markets performs an indispensable function by absorbing surpluses of capital.[28] For multinational corporations today, the whole world thus takes the form of a "profitable oyster".[29]

Capitalism's general tendency to reduce turnover time and obliterate geographical hindrances to accumulation means that it systematically compresses space and time. Distance is typically measured in time. As the time it requires to connect distinct geographical points shortens, space is "annihilated": the world of the high-speed jet, electronic communication, and instantaneous computerised business exchange, is "smaller" and "faster" phenomenologically than that of the stage coach and pony express. Capitalism increasingly makes the experiences of simultaneity and instantaneousness definitive features of modern existence. However overused and potentially misleading, the image of the "global village" does capture constitutive features of the phenomenological horizons of contemporary civilisation. If villagers in a rural community in Norway can experience the same thing at the same time (for example, a financial transaction) as city-dwellers in Toronto, or a taxi driver in Mexico City watches, via television, as police fire on protestors in Nigeria, a synchronisation of time and space occurs. Of course, neither time nor space have literally been compressed. But the "shrinkage" and concomitant "speed-up" in the pace of economic processes and social life at large, motored by capitalist development,

[26] David Harvey, *Justice, Nature & the Geography of Difference* (Oxford: Blackwell Publishers, 1996), pp. 240–1.
[27] Harvey, *The Condition of Postmodernity*, n. 25 above, p. 240.
[28] Harvey, *Justice, Nature & the Geography of Difference*, n. 26 above, p. 241.
[29] Ibid., p. 246.

does seem to bring "there" closer to "here" than it once was, while concurrently providing us with a sense that the "future" is always poised to collapse into the "present".[30]

On one level, there is nothing novel about the contemporary experience of time and space compression. As the cultural historian Stephen Kern shows in a path-breaking study, many of the great achievements of philosophy, literature, and art in the last century can be effectively interpreted as attempts to grapple with the compression of space and time that has always played an important role in industrial capitalism.[31] On another level, it would be mistaken to ignore the ways in which recent technological innovations —most important perhaps, computerisation— have suddenly deepened our consciousness of instantaneousness and simultaneity.[32] Although always an essential feature of modern capitalism, the compression of space and time has been especially intense in recent decades, in part as a consequence of the international capitalist economic crisis of the early 1970s which ultimately generated major innovations in production and consumption.[33] The Internet, satellite communication, lasers, information processing, and transistors: each has contributed unambiguously to alterations in the time and space parameters of economic action.

In my view, the capitalist-induced compression of space and time contains profound implications for all three sources of the elective affinity between capitalism and the rule of law described above. Even if a kinship between capitalism and a relatively formalistic model of the rule of law obtained throughout much of the history of modern capitalism, the ongoing globalisation of the capitalist economy suggests that it no longer does. Economic globalisation requires political and legal supports. Nevertheless, the political and legal infrastructure of globalisation bears little resemblance to the liberal model of the rule of law and its emphasis on the virtues of formal legality (generality, clarity, publicity,

[30] Though in different ways, the idea of a compression of space and time plays a role in the work of many recent theorists of globalisation, including Anthony Giddens and Paul Virilio: Anthony Giddens, *The Consequences of Modernity* (Stanford: Stanford University Press, 1990); Paul Virilio, *Speed and Politics* (New York: Semiotext, 1986). For now, I am unable to examine the pros and cons of distinct formulations of this concept; my main concern here is its potential value as a conceptual instrument for making sense of ongoing recent legal trends. It is important to note that a consensus seems to exist that capitalism is a major driving force behind the compression of space and time. But it may not be the only driving force.

[31] Stephen Kern, *The Culture of Time and Space, 1880–1918* (Cambridge: Harvard University Press, 1983).

[32] Of course, this consciousness also varies according to one's location in the social structure. A Wall Street currency trader, operating by means of high-speed electronic communication, is likely to have a different sense of time and space than, for example, the newspaper vendor, stationed daily at a street corner outside the currency trader's office, desperately trying to get rid of his daily allotment of newspapers. The manner in which social position and time-consciousness are interrelated is an important theme in Georges Gurvitch, *The Spectrum of Social Time* (Dordrecht, Holland: Reidel, 1964).

[33] Harvey, *Justice, Nature & the Geography of Distance*, n. 26 above, p. 245. I address some of the features of these innovations below.

prospectiveness, and stability in rules).[34] I start with a discussion of the immediate legal implications of the most recent bout of time and space compression (1), before examining its impact on the protective (2) and calculative (3) functions of the rule of law.

1. Time and space compression: legal implications

Recall from our discussion above that the rule of law in part traditionally served as an instrument for reducing insecurities stemming from the distance and duration of economic activity. Yet the capitalist-induced compression of space and time means that some of those uncertainties now have already been dramatically reduced, especially in economic sectors in which transactions increasingly rely on high-speed information, communication, and transportation technologies. Advanced technology takes care of at least some of the functions performed in an earlier phase in the history of capitalism by a liberal legal code consisting of clear, general, and relatively stable norms. Take, for example, the case of a present-day currency trader on Wall Street. Merely by pressing a few keys on her computer, vast amounts of currency are immediately exchanged within a few brief moments. Does our Wall Street currency trader experience the same need for a robust rule of law as her historical predecessors? Computerisation has drastically reduced the uncertainty resulting from the duration and distance of her business; she is likely to consider her eighteenth century predecessor's old-fashioned insistence on the virtues of a relatively unchanging, stable legal code quaint. Rapid-fire transactions with her peers in Frankfurt or Singapore are unlikely to be faultily disrupted even by quick changes in the law because her transactions are computerised, whereas alterations in the law, thus far, are not. She may find rapid changes in the law inconvenient, but hardly the immediate peril that it was to an eighteenth century merchant whose business relied on long and risky voyages.

This example is intended as more than an academic thought-experiment. As a matter of fact, the legal substructure of international finance and banking is remarkably underdeveloped, characterised for the most part by a set of recommended "best practices" exhibiting little formality or clarity. The main sites for the regulation of international banking (the Basel-based Bank for International Settlements and the Committee of Banking Supervisors) thus far have shown a striking preference for open-ended, flexible guidelines, in part because of the diversity of banking practices worldwide, but also in deference to the

[34] For some of the details: Saskia Sassen, *Losing Control?: Sovereignty in an Age of Globalization* (New York: Columbia University Press, 1996); William E. Scheuerman, "Economic Globalization and the Rule of Law" (1999) 6 *Constellations* forthcoming; Gunther Teubner (ed.), *Global Law Without a State* (Aldershot: Dartmouth, 1997).

free-wheeling and irresponsible forms of speculative "casino capitalism" now commonplace in the financial world.[35]

Even though a growing number of politicians are now expressing reservations about the dangers of the present system, international bankers have hardly been crying out for a system of transnational coordination based on clear, general, and stable state-backed norms. To the contrary, they generally have greeted attempts to develop such a system with scepticism and even hostility. If my analysis here is correct, there is at least one reason for this surprising challenge to the traditional view of an elective affinity between capitalism and the rule of law: the compression of space and time reduces the dependence of certain entrepreneurs on classical liberal law. Particularly within the financial and capital markets, simultaneity and instantaneousness function to reduce the economic agent's reliance on the rule of law as an instrument for counteracting uncertainty stemming from the duration and distance of commercial life, and thus it is no surprise that legal trends there conflict so dramatically with the traditional liberal model.

To be sure, computerised currency trading is an extreme example within the international economy, since simultaneity and instantaneousness play an especially prominent role there.[36] By the same token, the ongoing compression of space and time is likely to affect an increasingly broad range of commercial activities, as technological innovations diminish economic insecurity stemming from the duration and distance of economic exchange. Legal trends presently visible within the financial and banking sectors —most important, a preference among economic actors for porous, open-ended law— may represent a foreshadowing of the basic contours of an increasingly significant range of legal arenas.[37]

2. Impact on the protective functions of the rule of law

What then of the rule of law's classical protective functions? Notwithstanding its obvious strengths, the traditional liberal view here obfuscates the ways in which the ongoing globalisation of economic life potentially renders some of its core elements anachronistic as well.

[35] Tomaso Padoa-Schioppa and Fabrizio Saccomanni, "Managing a Market-Led Global Financial System", in Peter B. Kenen (ed.), *Managing the World Economy: Fifty Years After Bretton Woods* (Washington, D.C.: Institute for International Economics, 1994), pp. 259–60; Hirst and Thompson, *Globalization in Question*, n. 4 above, pp. 130–6. On speculative trends within international banking and finance, Susan Strange, *Casino Capitalism* (Oxford: Blackwell Publishers, 1986).

[36] In this vein, it is also striking that the legal substructure of "cyberspace commerce" —that is, business conducted via the Internet— remains badly underdeveloped. For a description of some of its legal weaknesses: Debora Spar and Jeffrey Bussgang, "Ruling the Net" (1996) 74 *Harvard Business Review* 125–33.

[37] For a vivid picture of changes in business practice already underway as a result of computerisation: Philip B. Evans and Thomas R. Wurster, "Strategy and the New Economics of Information" (1997) 75 *Harvard Business Review* 71–82.

Perhaps the most immediate institutional manifestation of capitalism's drive to overcome the limitations of space and time is the ascent of the mammoth multinational corporation (MNC), operating simultaneously in many parts of the world, and capable of exchanging goods and services across national borders at great speed. One way in which capitalism transforms the space and time horizons of economic activity is simply by generating "bigger" and "faster" firms. According to most accounts, globalisation is deepening economic concentration in the world economy, in part because of the exorbitant start-up costs entailed by advanced technology. Oligopolistic tendencies are more pervasive within the global economy than generally acknowledged by neo-liberal rhetoric or theory.[38] Increasingly in possession of more economic muscle than all but the richest members of the international state system, MNCs are also managing to outfit themselves with legal authority rivalling that of the nation-state itself. Many of the substantive norms of international business law are directly determined by the huge "industry leaders" who dominate the market, and a growing number of interstate economic agreements point in the direction of placing private businesses and nation-states on a level playing field in terms of legal status. States soon may no longer be the sole bodies in possession of legally recognised sovereign power within the international order. More and more, corporations exercise "sovereign" powers of law-making, while possessing legal "rights" no less impressive than those of the nation-state, the traditional carrier of sovereignty within modern times.[39]

Given the growing importance of MNCs to economic life, an ever more impressive range of economic activities are necessarily intrafirm in character. Yet intrafirm decision-making structures exhibit "more flexibility, and . . . less

[38] John Stopford and Susan Strange, *Rival States, Rival Firms: Competition for World Market Shares* (Cambridge: Cambridge University Press, 1991), pp. 6–72, 92–7; Robert Gilpin, *The Political Economy of International Relations* (Princeton: Princeton University Press, 1987), pp. 215–16. As I write this, a leading German daily reports that worldwide purchases of foreign firms and mergers broke all previous records in 1997. "Euro läßt Fusionen rollen. Konzentrationswelle erreicht weltweit einen Höchststand" *Frankfurter Rundschau* (15 September 1998), p. 14.

[39] Industry leaders within international marine insurance, for example, are the main source of international law in their sphere, and large firms played a decisive role in the standardisation of international sales terms. Of course, "home states are increasingly prepared to formulate both domestic and economic policy with the interests of their home-based MNCs in mind", which often amounnts to assuring legal mechanisms tailored to suit their preferences. Peter Muchlinski, "'Global Bukowina' Examined: Viewing the Multinational Enterprise as a Transnational Law-Making Community", in Teubner, *Global Law Without a State*, n. 34 above, pp. 85–9, 91. NAFTA grants firms rights hitherto generally limited to nation-states: Chapter 11(B) allows private businesses to submit complaints against member states to a three-member tribunal; member states and private firms have equal rights (!) to name members to the tribunal. The latest draft of the OECD Multilateral Agreement on Investments (MAI) points in a similar direction. For a defence of the trend towards outfitting corporations with rights thus far only possessed by nation-states: DeAnne Julius, "International Direct Investment: Strengthening the Policy Regime", in Kenen, *Managing the World Economy*, n. 35 above, pp. 282–3. Muchlinski criticises this tendency, and I have also done so in: Scheuerman, "Economic Globalization and the Rule of Law", n. 34 above. Political scientists and jurists have been slow to grasp the dangers entailed by placing corporations on the same legal playing field as nation-states. Political activists and intellectuals have been more adept at doing so: Noam Chomsky, "Power in the Global Arena" (1998) 230 *New Left Review* 23–7.

regard for decision-making consistency that might be acceptable for legal relations".[40] Profit-making and consistent norms and procedures hardly go hand-in-hand. For example, MNCs typically prefer differentiated labour standards and wage scales, for the familiar reason that it remains economically advantageous to treat workers in poor regions with less respect than those in the rich countries. Although contractual relations may formally exist between distinct component units of an MNC, such contracts often are no more than managerial orders written up in legal form, in reality lacking the minimal attributes of the classical consent-based free contract.[41] P.S. Atiyah's prescient observation that classical liberal forms of contract have declined in our century in part as a result of the increasingly prominent role of massive corporations, in which "relationships are conducted by administrative procedures and not by market contracts", takes on heightened significance given the growing role of massive "global players" in the world economy.[42] In short, even if we ignore the obvious institutional differences between privately-owned MNCs and nation-states, it is difficult to claim that the internal structure of MNCs displays even minimal features of legality. As MNCs take on an ever more prominent place within the world economy, classical liberal legal forms once essential to interfirm exchanges between economic competitors are thereby deprived of much of their previous significance as well.

The problem of intrafirm economic relations is only the tip of the iceberg. Few countries today can afford to brush off the prospect of foreign investment, and the decisive role of the MNCs within the world economy (in particular, their virtual monopoly on advanced technology) generates fierce competition among nation-states forced to bend over backwards in order to attract and keep foreign business. It is widely-acknowledged that the relative mobility and vast size of the MNCs often allows them to neutralise the regulatory capacities of the existing nation-state. Less well-known is that non-classical legal forms typically serve as an indispensable instrument for MNCs in their quest to do so. The structural advantages enjoyed by the MNCs in the international political economy permits them to turn the tables on the nation-state and its once impressive arsenal of administrative instruments. Big capitalist enterprise's reliance on the classical protective functions of the rule of law declines as well.

Two examples of this trend have to suffice for now. MNCs are notoriously undertaxed today, in part because they effectively exploit discrepancies among and between national tax codes. International business taxation exhibits extraordinary unevenness and disorder, and MNCs have for the most part employed every means to keep it that way. The reason for this is obvious enough: MNCs "legally" evade paying taxes by taking advantage of loopholes and gaps deriving from the complex and inconsistent structure of international business taxation, and they often prefer ad hoc, closed-door negotiations conducted by

[40] Muchlinski, "'Global Bukowina' Examined: Viewing the Multinational Enterprise as a Transnational Law-Making Community", n. 39 above, p. 82.

[41] Ibid., p. 83.

[42] P.S. Atiyah, *The Rise and Fall of Freedom of Contract*, n. 3 above, p. 724.

corporate tax lawyers and government officials which determine how much they are to pay.[43] In a global economic context in which (1) MNCs often possess unprecedented mobility, and (2) attracting MNC investment seems essential to economic well-being, corporate representatives possess obvious structural advantages when engaging in ad hoc negotiations with tax officials even from rich countries. In a similar manner, MNCs (with political support from their home countries) thus far have managed to ward off even modest attempts to set up enforceable international legal codes promulgating proper forms of business conduct. Recent United Nations-sponsored plans to challenge an onerous history of MNC-meddling in the internal affairs of host countries have been beaten back, and cautious attempts by the International Chamber of Commerce and OECD —hardly principled critics of international business— have resulted in recommendations possessing at best the character of "soft" law.[44] In this sphere as well, MNCs for the most part seem to prefer an international regulatory scenario plagued by enormous inconsistencies and discrepancies, since its inchoate structure provides loopholes galore for maximising the privileges enjoyed by the MNC within the global economy.

MNCs also have less to fear from the possibility of arbitrary government than small and medium-sized firms lacking their vast resources and mobility. The mere spectre of "unfair" or "discriminatory" treatment by a host country often suffices as a disincentive to invest there in the first place. And even after the decision has been made to invest in a particular locality, MNCs may exhibit few reservations about closing down and moving elsewhere at signs of a worsening "investment climate" —for example, when a newly elected government threatens to make MNCs pay their fair share of taxes in accordance with standing local regulations. Furthermore, because especially small and medium-sized states increasingly depend on MNC investment for economic development, their interest in undertaking hostile forms of "arbitrary" action against foreign capital is substantially reduced from the outset. On the contrary, the importance of MNC investment to economic development means that prospective host states typically compete to attract foreign investment by promising what in effect amount to special rights and privileges (tax breaks, direct and indirect subsidies, government outlays for research).[45] Accordingly, anxieties experienced by MNCs today stem not from the traditional spectre of discretionary or

[43] Sol Picciotto, *International Business Taxation: A Study in the Internationalization of Business Regulation* (London: Weidenfeld & Nicolson, 1992).

[44] Muchlinski, "'Global Bukowina' Examined: Viewing the Multinational Enterprise as a Transnational Law-Making Community", n. 39 above, pp. 90–6; Bernhard Grossfeld, "Multinationale Unternehmer als Anstoß zur Internationalisierung des Wirtschaftsrechts", (1980) 32 *Wirtschaft und Recht*; Norbert Horn, "Die Entwicklung des internationalen Wirtschaftsrechts durch Verhaltensnormen", (1980) 44 *Rabels Zeitschrift für ausländisches und internationales Privatrecht*.

[45] This is the "new pragmatism" between host states and foreign capital described by Stopford and Strange in *Rival States, Rival Firms: Competition for World Market Shares*. The World Trade Organization's record of success in warding off practices of this type is mixed: Ernest H. Preeg, *Traders in a Brave New World: The Uruguay Round and the Future of the International Trading System* (Chicago: University of Chicago Press, 1995).

arbitrary state intervention hostile to the pursuit of profits, but instead from the prospect that the special incentives they enjoy may be less generous than those gained by competitors operating elsewhere. From the perspective of the MNCs, government "discretion" is often a problem only to the extent that it may be inadequately lucrative.

This situation is clearly distinct from that envisioned by classical liberal theorists who wrote so convincingly of the dangers of political arbitrariness to commercial life. For sure, these perils remain real for segments of the business community unable to neutralise administrative authority by "playing off" nation-states against each other. Yet for an increasingly significant sector of the international economy, the classical protective functions of the rule of law lack the overriding sigificance they once possessed. If Czech authorities fail to enforce laws requiring the protection of a struggling "mom-and-pop" grocery store unable to pay protection money to the "Mafia", the store probably will go under. In contrast, if Coca-Cola gains word of recalcitrant officials who refuse to enforce anti-"Mafia" laws, it may threaten Czech authorities with the prospect of opting for another location for its next plant. On the basis of such threats, Coca-Cola not only will be able to make sure that anti-"Mafia" laws are enforced, but may be able to garner an array of additional lucrative privileges. Of course, similar threats from the local grocery-store are likely to gain nothing but a dismissive glance from a lower level civil servant.

When early liberals like Locke and Montesquieu first described the protective functions of the rule of law, capitalist enterprise for the most part was small-scale and tied to a specific geographical location; from the perspective of the fledgling entrepreneur, the political authorities who exercised sovereignty within that locality were a force to be reckoned with. At the end of the twentieth century, the scope and scale of capitalist production reduces both the de facto and de jure significance of the sovereign power of the nation-state and thereby simultaneously diminishes the importance of the traditional protective functions of the rule of law for the largest and most mobile units of capital. The main problem posed by globalisation is less that international business can only preserve its autonomy by limiting state power by means of the rule of law, than that the democratic nation-state can only hope to maintain its independence in relation to international capital by counteracting the virtually universal competitive rush to provide foreign firms with special rights and privileges. Needless to say, this competition contains worrisome implications for the regulatory capacity of the democratic state, let alone its ability to achieve a necessary minimum of social cohesion: the rush to provide foreign capital with investment incentives contributes to a "race to the bottom" in which states embrace a now-familiar coterie of neo-liberal reforms, including corporate tax cuts, the roll-back of the welfare state, and relaxed social and environmental regulations. Whether or not liberal democracy will still be able to manage the negative side-effects of globalisation effectively, let alone assure the social consensus essential to its stability, remains an open question.

In an astonishing historical reversal, discretionary authority for the sake of international business now threatens the democratic nation-state, whereas international business often gains directly from non-general, irregular regulations (for example, special tax cuts for foreign investors). States need increased generality, consistency, and stability within international regulation in order to preserve their integrity, whereas privileged international economic interests seem perfectly happy with legal inconsistency and irregularity among states forced to court them.[46]

3. Impact on the calculative functions of the rule of law

The capitalist-induced compression of space and time also raises difficult questions in reference to the calculative ethos of capitalist enterprise which Weber and many influenced by him considered so important for its kinship to the rule of law. Above I alluded to the fact that Weber's picture of the capitalist entrepeneur has long been subject to a series of scathing criticisms. Joseph Schumpeter, for example, early on suggested that Weber's model obscured the core of capitalist entrepreneurship, namely the ability to act in unforeseen ways that often seem irrational from the perspective of pre-existing forms of economic behaviour. Weber's exaggerated focus on the calculative ethos of the capitalist entrepreneur allegedly stumbles because "[t]he nature of the innovation process, the drastic departure from existing routines, is inherently one that cannot be reduced to mere calculation, although subsequent imitation of the innovation, once accomplished, can be so reduced".[47] For Schumpeter, the *differentia specifica* of capitalist entrepreneurship, namely the capacity to pursue economic innovation by piercing the crust of worn-out commercial routine, is poorly captured by a model of capitalism in which predictability and calculability are described as its dominant principles. For Schumpeter, the classical entrepreneur is an heroic and even charismatic figure precisely because he shatters predictable and calculable modes of economic activity.

This is not the place to take sides in one of the great debates in twentieth century economic theory. Yet recalling Schumpeter's critique of Weber at least brings attention to the fact that one influential statement of the idea of an elective affinity between capitalism and the rule of law relies, at least to some extent, on a controversial model of the capitalist enterprise. What then is the status of

[46] By no means does this reversal render the protective functions of the rule of law wholly anachronistic. My point here is simply that it does so for the most privileged segments of the corporate world.

[47] Nathan Rosenberg, "Joseph Schumpeter: Radical Economist", in Mark Perlman and Yuichi Shionoyo (eds.), *Schumpeter in the History of Ideas* (Ann Arbor: University of Michigan Press, 1994), p. 48. For textual support for this reading: Joseph Schumpeter, *The Theory of Economic Development* (Cambridge: Harvard University Press, 1934), pp. 79–83. Schumpeter is critical of Weber's "protestant ethic thesis" in Schumpeter, *Business Cycles* (New York: McGraw Hill, 1939), p. 228.

the calculative ethos described by Weber in the context of economic globalisation? In an important critical discussion of this question, Harvey argues that a recent bout of innovation in information, communication, and transportation technologies has generated far-reaching shifts in the workings of the contemporary capitalist enterprise. Novel economic possibilities provided by new technologies place a special "premium on 'smart' and innovative entrepreneurship, aided and abetted by all the accoutrements of swift, decisive, and well-informed decision-making".[48] Heightened possibilities for simultaneity and instantaneousness make forms of "flexibility with respect to labour processes, labour markets, products, and patterns of consumption" possible on a scale that would have stunned earlier generations of entrepreneurs.[49] Successful capitalist enterprises today are characterised by their prowess at rapidly adjusting to new information and new techniques, resulting in "greatly intensified rates of commercial, technological, and organisational innovation".[50] The compression of space and time not only provides increased opportunities for flexibility and mobility, but the successful capitalist knows how to exploit them.

We would probably be well advised to take Harvey's model of entrepreneurship as describing trends within the global economy, but hardly the whole story; it would be a mistake to exaggerate the immediate impact of new technologies on the capitalist enterprise. Nonetheless, his account provides a helpful starting point for making sense of a number of recent legal trends.

For example, Atiyah refers to "the pace of change in modern society" as one of the sources for the decline of classical forms of free contract.[51] As Atiyah notes, contracts in our century, in distinction to their nineteenth century predecessors, tend to provide ample possibilities for parties engaged in an economic exchange to renegotiate their agreement "on terms which are open to continuous adjustment as long as the relationship lasts".[52] Possibilities for flexibility are now built in to the structure of contracts, suggesting to Atiyah that a "growing recognition that the opportunity to change one's mind is a valuable right" has played an important role in the transformation of freedom of contract.[53] My argument here about the capitalist-induced compression of space and time places Atiyah's account in a fresh light. The acknowledgement of a right "to change one's mind" surely in part is motored by ongoing changes in the structure of capitalist enterprise, according to which flexibility and the possibility of rapid-fire adjustments become decisive to economic success. Contracts allow for

[48] Harvey, *The Condition of Postmodernity*, n. 25 above, p. 157.
[49] Ibid., p. 147.
[50] Ibid.
[51] Atiyah, *The Rise and Fall of Freedom of Contract*, n. 3 above, p. 713.
[52] Ibid., p.717. For example, "[t]he modern commercial transaction is, in practice, apt to include provision for varying the terms of exchange to suit the conditions applicable at the time of performance. Goods ordered for future delivery are likely to be supplied at prices ruling at the time of delivery; rise and fall clauses in building or construction works are the rule and not the exception; currency-variation clauses may well be included in international transactions": ibid., p. 714.
[53] Ibid., p. 756.

economic actors to change their minds in part because the right to change one's mind is ever more indispensable in an economy characterised by the growing importance of exploiting opportunities provided by instantaneousness and simultaneity. Atiyah obscures this feature of the story because he chiefly attributes the decay of classical "executory" contracts to the rise of the welfare state and growth of public enterprise. Though accurate, this interpretation fails to explain sufficiently why flexibility within contract law is increasingly common in areas of law in which neither public enterprise nor the welfare state play an important role.[54]

According to most accounts, flexibility is especially widespread within international economic law.[55] Given my argument here, this is unsurprising as well: "global players" thus far have proven most adept at making use of new information, communication, and transporation technologies, and they come closest to fitting the model of the contemporary capitalist enterprise described by Harvey. Correspondingly, they seem most comfortable with legal institutions providing generous possibilities for flexible modes of conflict resolution, particularly when decision-makers sympathise with the basic ideological orientation of the international business community as a whole.

For example, analysts of the burgeoning field of international business arbitration generally agree that its legal structure exhibits little formality and consistency. Nonetheless, a popular "how-to" literature tailored for the international business community praises arbitration as superior to costly, time-consuming, and purportedly rigid forms of "bureaucratic" adjudication, and international business is opting, in ever greater numbers, to resolve conflicts by means of arbitration services over traditional forms of court-based adjudication.[56] Traditional liberal law, it seems, has become too inflexible and unwieldy for global entrepreneurs hoping to make optimal use of the economic possibilities provided by the compression of space and time. In a world in which economic success requires speedy reactions to complex, fast-moving economic shifts, a system of legal coordination offering substantial opportunities for discretionary decision-making contains some obvious advantages for entrepreneurs in need of multiple chances to "change their minds". Economic transactions today take place at a fast pace; sudden changes in market conditions demand flexibility and fast reactions. Legal forms permitting flexible decision-making potentially provide a framework in which economic actors can adjust effectively to the ever-changing dynamics of the market-place.

In contrast, classical modes of stable, general law may appear to represent an impediment to the rapid-fire responses required by the ever-changing dictates of

[54] For some examples: ibid., p. 759.

[55] For an important survey, Teubner, "'Global Bukowina': Legal Pluralism in the World Society", in *Global Law Without a State*, n. 39 above, p. 3.

[56] See the essays collected in Thomas Carbonneau, (ed.), *Lex Mercatoria and Arbitration* (Dobbs Ferry, NY: Transnational Juris Publications, 1990). Also Jeswald W. Salacuse, *International erfolgreich handeln* (Frankfurt a.M.: Campus Verlag, 1992), pp. 111–38.

the international economy. Traditional codified law may seem static and even lifeless, a leftover from a past poorly suited to the latest imperatives of an economy in which the ability to adapt rapidly separates winners from losers.

III. POLITICAL LESSONS

Capitalism's once intimate relationship to the rule of law seems ever more distant and estranged. The elective affinity described by so many legal and political theorists on both the left and right belongs, for the most part, to the trash can of legal and intellectual history. Although it remains true that every functioning capitalist economy requires some minimum of legal protections (private property, a system of contract), even that minimum is more pliable than generally acknowledged.[57] By no means can we legitimately endorse the view that capitalism and a robust rule of law, based on a system of clear, general, stable, prospective, public norms, are likely to go hand-in-hand. On the contrary, economic globalisation flourishes precisely where such legal forms are lacking.

Despite this surprising historical shift, the rule of law remains essential to any worthwhile democratic polity. Clear, general, stable, prospective, public norms are necessary if we are to tame arbitrary power and prevent tyranny. The fact that the rule of law is no longer essential to global capital hardly makes it any less important as a basic assurance of legal security for democratic citizens. In addition, I have tried to suggest here that the rule of law today is gaining in utility as a protection for democratic polities subject to the whims of giant economic interests; discretionary decision-making, for the benefit of international business, poses a growing threat to the political effectiveness of the democratic state. If we are to ward off the emerging spectre of an international economic tyranny, in which mammoth economic interests employ their structural advantages to dictate economic and social policy to the rest of us, we immediately need to move towards the establishment of enforceable transnational economic, social, labour, health, and environmental standards; in this way as well, the rule of law can continue to perform basic protective functions. At the very least, such standards can help minimize the dangers of social and environmental "dumping" and thereby counteract the disastrous trend among nation-states to engage in ruthless competition to attract foreign investment. If they are to prove effective, these standards are going to have to take on many of the attributes of legality described by classical liberal theory: in light of the structural advantages enjoyed today by international business, it often remains best positioned to take advantage of ambiguities and discrepancies within legal and regulatory standards. In an earlier essay, I noted that vague types of deformalised law within economic and social policy often are exploited by the biggest and best organised

[57] On contracts: Atiyah, *The Rise and Fall of Freedom of Contract*, n. 3 above. On property: Renner, *Die Rechtsinstitute des Privatrechts und ihre soziale Funktionen*, n. 22 above.

interests within the domestic political economy.[58] If I am not mistaken, these perils are even greater in the global arena, where political mechanisms capable of counteracting the potential dangers of non-formal law are either under developed or non-existent. In the context of a democratic polity in which modest and lower-income social constituencies have been able to establish a measure of political influence, at least some of the dangers of non-formal law can be reduced.[59] Within the contemporary global political economy, however, it remains unclear what present-day institutions might perform a similar function.

But do not vast cultural, ethical, and religious differences within the global arena make it impossible to establish a strengthened set of binding transnational economic regulations? Radical critics of the rule of law long have argued that modern moral and political pluralism necessarily undermines the traditional project of generating legal determinacy by means of a liberal model of the rule of law.[60] Is not this problem likely to prove even more severe in the international arena?

Of course, the challenges posed by the "fact of pluralism" are greater on the international than on the domestic stage. However, we would probably be well advised not to exaggerate them. Effective rules are often achievable even in the context of far-reaching moral and political disagreements, and general rules remain an indispensable instrument for facilitating social cooperation in light of real-life limits of time and energy that are likely to prevent legal and political actors from solving fundamental moral conflicts. As Cass Sunstein has noted in a different context:

> "[p]eople can urge a 60-mile-per-hour speed limit, a prohibition on bringing elephants into restaurants, a ten-year maximum sentence for attempted rape, and much more without taking a stand on debates between Kantians and utilitarians . . . [R]ules sharply diminish the level of disagreement among people who are subject to them and among people who must interpret and apply them. When rules are in place, high-level theories need not be invoked in order for us to know what rules mean, and whether they are binding".[61]

In short, there is no principled reason why pluralism makes it impossible for nation-states to cooperate in establishing stricter transnational social and environmental standards, as well as a more formalistic, transparent international economic code.

The only answer to the crisis of the rule of law is thus more of the rule of law —operating on a transnational scale. The relative autonomy of the democratic

[58] Scheuerman, "The Rule of Law and the Welfare State: Towards a New Synthesis", n. 7 above.

[59] Franz L. Neumann made this point over sixty years ago in "The Change in the Function of Law in Modern Society" (1936), in William E. Scheuerman (ed.), *The Rule of Law Under Siege: Colllected Essays of Franz L. Neumann and Otto Kirchheimer* (Berkeley: University of California Press, 1996), pp. 101–41.

[60] In this spirit, see Roberto Mungabeira Unger, "The Critical Legal Studies Movement" (1983) 96 *Harvard Law Review* 571.

[61] Cass Sunstein, *Legal Reasoning and Political Conflict* (New York: Oxford University Press, 1996), pp. 110–11.

polity in the face of international business can be preserved, but only if democratic states cooperate closely to establish new modes of rigorous transnational regulation.[62]

Needless to say, the struggle for strict transnational norms is sure to face fierce opposition from both privileged economic interests and wealthy nation-states which have the most to gain from the status quo. Yet those of us committed to liberalism's most important weapon against arbitrary power, the rule of law, have no choice but to throw ourselves into the political battle at hand. I hope that I have made an initial contribution to that battle here by discrediting outdated illusions about the purported kinship between capitalism and the rule of law.

When writing about the deleterious impact of laissez-faire ideas on modern liberal democracy, John Dewey commented that "[t]raditional ideas are more than irrelevant. They are an encumbrance".[63] This is true of the traditional notion of an elective affinity between capitalism and the rule of law as well.

[62] For one discussion of this prospect: David Held, *Democracy and the Global Order: From the Modern State to Cosmopolitan Governance* (Stanford: Stanford University Press, 1995).

[63] Cited in John Patrick Diggins, *The Promise of Pragmatism: Modernism and the Crisis of Knowledge and Authority* (Chicago: University of Chicago Press, 1994), p. 264.

13

Supranational Challenges to the Rule of Law: The Case of the European Union

JOHN P. McCORMICK

Today, an increasingly substantial amount of legal activity that directly affects the lives of people around the world functions at levels beyond the traditional boundaries of nation-states. The outcomes decided, for instance, in the World Trade Organization (WTO) or the North American Free Trade Agreement (NAFTA) dispute panels, or in the European Court of Justice (ECJ), affect individuals today as profoundly as did decisions rendered in constitutional or lower-level national courts just a generation ago. One question facing legal scholars today is whether judicial activity at these supranational levels can be understood to conform with principles of the rule of law. In an attempt to begin answering the question of the status of law in the contemporary global arena, I propose to examine the extent to which the rule of law operates within the context of one region of the globe where supranational organisation is most advanced, the European Union (EU). To do so, I focus on the contemporary role of the ECJ and EU law in the process of European integration.

I. THE LEGACY OF THE NATIONAL *RECHTSSTAAT*(S)

According to traditional notions of the rule of law, courts have been semi-autonomous institutions within nation-states. In the *Rechtsstaat* paradigm of the constitutional state, courts function as neither the puppets of legislatures who make laws, nor the ignored-as-irrelevant weaker institutional siblings of executive authorities. Rather, courts are granted powers of oversight over both of these ostensibly more powerful arms of government. While to some extent bound by the terms of statutes produced by parliamentary majorities in adjudication, constitutional courts in many contexts have had the authority to strike down such legislation on the basis of constitutional principles. Moreover, while possessing no specific enforcement powers of their own, courts have been served with more or less good faith by the effective enforcement of their judicial

decisions by executive actors. Thus a functioning separation of powers is an institutional *sine qua non* of the traditional *Rechtsstaat*.[1]

In other words, proper court activity under the rule of law in this admittedly highly abstract model entails a certain balance between autonomy from, and dependence on, other institutional actors. The ramifications of this limited-autonomy or enabled-constraint, depending on how one looks at it, is an institutional safeguard against the more political branches of government who might encroach upon the basic constitutional principles of the regime itself, such as protection of minorities, equality before the law, state non-intervention into spheres of economic freedom, the prohibition against *ex post facto* judgment, the relatively consistent reliance on precedent, etc.[2]

Along with the institutional requirement of the separation of powers, the traditional *Rechtsstaat* presupposes a sociological separation of state and society. State activity is confined to the guarantee of internal and external security of citizens, with the assumption that a self-regulating market economy will provide the substantive means by which all or at least the vast majority of citizens could achieve a good life.[3] The *Rechtsstaat* separation of state and society ensures the rule of procedurally legitimated law rather than the rule of executive-executed force. The executive arms merely concretise in specific circumstances the general contents of statutes; they serve to make real the semantic formulas of abstract and general norms.[4] Thus the grammar of legal rules itself has a distinctive character under the *Rechtsstaat*: abstract, general, formal and *conditional*. The latter characteristic is best illustrated by the "if x, then y" formulation of law under the *Rechtsstaat*. The structure of law itself limits state activity: the state may take appropriate action y only "if" certain circumstances x arise in social reality.[5]

Thus the delicate institutional position of the judiciary in the *Rechtsstaat* is preserved by the form of the kind of laws legislated and adjudicated in this largely nineteenth century model. Confined to the adjudication of cases testing statutes dealing with criminal law and a narrow conception of property, courts were rarely perceived as illegitimately encroaching on the responsibilities of the other branches. Thus they were rarely treated in a hostile manner by the other branches who would threaten their autonomy. When they did challenge the activity of other state institutions they would do so on the basis of explicit and often written constitutional provisions that were not open to varied or controversial interpretation.

[1] See Jürgen Habermas, *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy* (Cambridge, MA: MIT Press, 1996), pp. 245–6, 431. I rely on this recently published volume in this section as a guide in my ensuing discussion of the types of *Rechtsstaat*.

[2] On general elements of the rule of law, see George P. Fletcher, *The Basic Concepts of Legal Thought* (Oxford: Oxford University Press, 1996).

[3] Habermas, *Between Facts and Norms*, n. 1 above, pp. 174–5.

[4] Ibid., pp. 188–9.

[5] Ibid., p. 208.

As parliaments begin enacting, and courts gave up resistance to, the legislation or management of an increasingly industrial economy in the twentieth century, we perceive the emergence of what has been called, often perjoratively, the social-democratic rule of law or *Sozialrechtsstaat*. Courts now adjudicate more substantively defined principles of liberty and equality transcending those narrowly associated with property rights. State activity in all branches of government becomes concerned with issues of redistributive taxation, the collective organisation of labour, worker safety, social insurance, discouragement of monopoly, public control of fiscal markets, etc.

Whereas so-called *classical* administration under the *Rechtsstaat* was reactive to actual and specific social events, functionally divided among branches, restrained in its approach to intervention, in contrast, *service* administration under the *Sozialrechtsstaat*, which provides public goods, infrastructure, planning and risk prevention to society, is necessarily future-oriented to often hypothetical situations and expansive in approach.[6] Law no longer looks like conditional propositions with universal claims but rather takes the form of special legislation, experimental temporary laws, broad regulative imperatives with uncertain prognoses, blanket clauses, all couched within indefinite statutory language.[7]

From the standpoint of a strict conception of the traditional *Rechtsstaat* that understands the rule of law as comprising only rights against the state, legal guarantees of public order, and prohibitions on the abuse of "economic liberty", all maintained through general and abstract law, the *Sozialrechtsstaat* appears to be a "corruption" of, or "upheaval" against, law itself.[8] The very nature of law might be compromised by recourse to less than clearly defined conditional phrases and declarative imperatives that enable broad policies, fundamentally infused with appeals to substantive principles; a "remoralization" or "deformalization" of law, if you will.[9] The separation of powers is likewise jeopardised by an unhinging of executive and judicial activity from the express will of the legislature, thus expanding their power *vis-à-vis* the latter.[10] In short, for some, the *Sozialrechtsstaat* obscures the clarity and mechanical application of *Rechtsstaat* law due to the vagueness of social law and the discretion of its enforcers and adjudication.[11]

So much of the discourse distinguishing the two paradigms of the rule of law serves to validate one or the other: the neo-conservative versus the social democratic visions of law, respectively. Here, I want to assume, following many progressive legal scholars, especially Jürgen Habermas, a compatibility and mutually enforcing relationship between the traditional *Rechtsstaat* and the

[6] Ibid., p. 431.
[7] Ibid.
[8] Ibid., pp. 245–6, 431.
[9] Ibid., p. 246.
[10] Ibid., pp. 246, 431.
[11] Ibid., pp. 188–9.

Sozialrechtsstaat—a mutuality institutionalising the indispensability of freedom and equality for each other.[12] Habermas identifies this as the internal relation of private and public autonomy.[13] I have elaborated this at greater length elsewhere.[14] Let it be sufficient to say here that Habermas's *Aufhebung* of what he calls bourgeois-formal and welfare state-material law privileges neither the formal character of law-construction and adjudication, on the one hand, nor the substantive content of economic equality, on the other, but rather a discourse-centered process that renders contemporary law constituent of and by both.[15]

Habermas's discourse theory of law has as its goal not just the protection of equal private rights nor the realisation of the principles of social justice, but the retention of both in the achievement of "an exclusive opinion- and will-formation in which free and equal citizens reach an understanding on which goals and norms lie in the equal intent of all".[16] The essence of law is not just the pursuit of economic gain nor the attainment of economic equality, but the realisation of an active—or rather interactive—pursuit of the conditions by which these may be expanded in non-pathological (i.e., capitalist exploitative or state-socialist repressive) means. Habermas proposes a strategy whereby law-makers and adjudicators explicitly invoke the principles and methods characteristic of either bourgeois formal law (the *Rechtsstaat*) or welfare-state materialised law (the *Sozialrechtsstaat*) depending on whether a particular case suggests liberty-allowing or equality-encouraging priorities. Habermas deems this a reflexive theory of law.[17]

Yet, despite the charges by largely neo-conservative critics against the jurisprudential indeterminacy and judicial usurpation that supposedly arises from the *Sozialrechtsstaat* or its combination with the *Rechtsstaat*,[18] the cohesion of a specifically national-territorial constitutional order itself shielded courts from excessive backlash and unfaithful enforcement of their decisions in Fordist nation-states of the post-war era. Almost two centuries of entrenched arrangements of the separation of powers in the liberal democracies of Europe and North America insulated courts from shocks caused by the transition to the

[12] For other considerations on the compatibility of the rule of law and the welfare state see: in the European context, Günther Teubner (ed.), *Dilemmas of Law and the Welfare State* (New York: de Gruyter, 1988), and Neil MacCormick, *Legal Right and Social Democracy: Essays in Legal and Political Philosophy* (Oxford: Clarendon Press, 1984); and in the American context, Cass R. Sunstein, *After the Rights Revolution: Reconceiving the Regulatory State* (Cambridge, MA: Harvard University Press, 1990); Sunstein, *Free Markets and Social Justice* (Oxford: Oxford University Press, 1997); Bruce Ackerman, *Reconstructing American Law* (Cambridge, MA: Harvard University Press, 1983); and Ackerman, *Social Justice and the Liberal State* (New Haven: Yale University Press, 1981).

[13] Habermas, *Between Facts and Norms*, n. 1 above, p. 135.

[14] John P. McCormick, "Habermas's Discourse Theory of Law: Bridging Anglo-American and Continental Legal Traditions" (1997) 60 *Modern Law Review* 734.

[15] Habermas, *Between Facts and Norms*, n. 1 above, p. 135.

[16] Ibid., p. 270.

[17] Ibid., pp. 393, 410.

[18] The classic articulation of this position is Friedrich A. Hayek, *Law, Legislation and Liberty* (Chicago: University of Chicago, 1978), 3 vols.

social from the capitalist *Rechtsstaat* that might otherwise have thrown courts from the historical vehicle of constitutional politics.

Recent scholarship suggests that standards of both the traditional and the social-democratic rule of law are being challenged and, to some extent, compromised by recent supranational developments associated with globalisation and regionalisation. It is argued that international legal fora are faced with a greater encroachment on their functioning by state and multinational actors than national courts have been by their domestic institutional counterparts; that the enforcement of court decisions are at the mercy of the potentially arbitrary enforcement patterns of authorities not directly linked to them institutionally through territorially-bound constitutions; and that the purview of supranational court jurisdiction will not soon be expanded to include substantive social and political concerns but rather will remain constrained to more narrow economic ones.[19] I take up these concerns in an analysis of the recent literature on law and the European Court of Justice in the European Union. A supranational realisation of Habermas's discourse theory of the rule of law in a continental constitutional state is not feasible any time soon.[20] There is, moreover, the serious question of whether the reflexive theory of law is feasible outside of a national context. However, I will proceed in my analysis of the literature on European law with the assumption that the tension between traditional *Rechtsstaat* and welfare-state *Sozialrechtsstaat* principles is a more desired situation than the absence of one, the other or both at an EU level.

II. THE STATUS OF THE RULE OF LAW IN THE EU

Law is unequivocally central to the development of the EU. While the law's influence had perhaps been overstated in the past,[21] today, even the most state-centric commentators marvel at the transnational force of European law. This section establishes the general autonomy and efficacy of the ECJ and EU law, thus suggesting that the latter satisfy traditional criteria of the *Rechtsstaat*. However, whether the rule of law at a European level can fully take the place of

[19] The following volumes lay out quite well the constraints on an expansion of the rule of law in the global arena, while nevertheless advocating the pursuit of this very possibility: Hans Kochler, *Democracy and the International Rule of Law: Propositions for an Alternative World Order* (Berlin: Springer Verlag, 1995); David Held, *Democracy and the Global Order: From the Modern State to Cosmopolitan Governance* (Stanford: Stanford University Press, 1995); and Günther Teubner (ed.), *Global Law Without a State* (London: Dartmouth Publishing, 1996).

[20] Although Habermas would suggest otherwise: see "The European Nation State—its Achievements and its Limits: On the Past and Future of Sovereignty and Citizenship" (1995) II *Europäische Zeitschrift für Recht, Philosophie und Informatik* 2; "Remarks on 'Does Europe Need a Constitution?'" (1995) 1 *European Law Journal* 303; and "Citizenship and National Identity," App. II to *Between Facts and Norms*, n. 1 above.

[21] For instance, A. W. Green, *Political Integration by Jurisprudence* (Leyden: Sijthoff, 1961); Cappelleti, et al. (eds.), *Integration Through Law* (Berlin: de Gruyter, 1986). An indispensable contemporary guide to the role of law in European integration is Neil MacCormick (ed.), *Constructing Legal Systems: "European Union" in Legal Theory* (Amsterdam: Kluwer, 1997).

law in liberal democratic nation-states or, more hopefully, further advance principles associated with *both Rechtsstaat* models remains an open question. I deal with debates on this issue in the subsequent sections.

The ECJ has become the EU institution perhaps most autonomous of member state influence, even if it is not completely free of it. It is also the one with which individuals within European member states have a semi-direct relationship. This is a relationship reminiscent of citizen-government relations, and not, like other forms of EU-citizen interactions, merely a relationship mediated through the offices of the heads of state, ranking ministers or appointments to the European Commission. Unlike the Commission, the Council of Ministers and the European Council,[22] the ECJ's policy adjudicating is least beholden to *direct* member state sanction. (The exact extent of the Court's susceptibility to influence by the member states more generally is one of the central questions at issue in this essay.) European citizens interact with the ECJ through more intimate channels—appeals through local courts—than other more distant member state organs.

The ECJ's policy purview may be less wide-ranging than that of the other EU institutions, and certainly less than the high courts of most nation-states (lacking explicit criminal, family, educational, health, and, to some extent, social and human rights jurisdiction). But the expanding spheres of economic integration (single market, environmental and consumer protection) are beginning to overlap with many, if not all, social welfare concerns.

Most observers are somewhat astounded by the extent to which national courts refer cases to the ECJ, and how often and faithfully the former abide by the Court's rulings upon return. Several jurisprudential principles have established the de jure supremacy of European law over that of member states: *direct applicability* establishes the immediate validity of EU law within member states without the necessity of subsequent member state measures (i.e., enabling or specifying acts); *direct effect* establishes the recognition by member state courts of European law's conference of rights, or imposition of obligations, on particular individuals; and *primacy*, while not mentioned in the treaties of the European Community, or more recently, European Union, is assumed in ECJ and member state court decisions and establishes the supremacy of EU law over member state law on specific issues, even those dealt with in member state constitutions.

In the 1990s, the ECJ ruled in favour of compensation for those adversely affected by EU negligence (*Francovich*, Cases 6/90 and 9/90, 1992 ECR [1990] p. 5357); the harmonisation of member state social security programmes to protect

[22] The Commission is a seventeen-member body composed of a president and one or two nominees from each member state, approved by the European Parliament for a five-year-renewable term. The Commission functions as the administration of the EU. The Council of Ministers is composed of the ranking member state ministers of whatever policy sphere is being dealt with at a given time: e.g., the foreign ministers, the energy ministers, etc. The European Council is the meeting of the respective heads of governments of the member states. The general institutional framework of the EU is quite ably explained in Neil Nugent, *The Politics and Government of the European Union* (Durham: Duke University Press, 1996).

migrant workers and intra-EU immigrants in numerous cases; struck down unequal pension-eligibility ages in the United Kingdom (*Barber*, Case 262/88, 1990 ECR [1990] p. 1889); and decided numerous cases furthering the single European market (e.g., *Cassis de Dijon*, Case 120/78 ECR [1979] p. 649). Previously, the most important ECJ cases dealt with interinstitutional relations: e.g., the Council must wait for reports from the European Parliament (EP) before enacting legislation (Case 138/79, 1980); or on the extent of the EP's budgetary powers (Cases 89, 104, 114–17, 125–29/85, 1988).

The ECJ cannot initiate action itself, but has significant power when asked to rule: it can order a member state to fulfil its obligations to other member states and individuals; it can void EU acts if they are not adopted according to procedures laid out by Community treaties; and it can set fines unlimited in amount for misbehaviour. Most frequently, the ECJ is asked by member state courts to rule on specific points of law, and the Council or Commission can seek the ECJ's opinion on international agreements. Thus the ECJ and European law seem to function according to criteria associated with both the traditional liberal and social *Rechtsstaat*, through the adherence to principles of legal supremacy, and the Court's role as arbiter of it; jurisdiction over the institutional relations of the other "branches" of EU governance; and the Court's role in managing the political economy of the Union. How extensively we may want to understand the Court as a traditional *Rechtsstaat* or *Sozialrechtsstaat* institution is open to vigorous debate.

III. EU LEGAL PROCESS AS ROBUST RULE OF LAW

The normative vision of scholars—mostly lawyers—with a generally optimistic view of the ECJ's power can be summed up as follows: the EU develops as a supranational legal order that may be appealed to by European citizens through local courts on issues initially economic, but increasingly social and political. This process facilitates a normative interaction between, on the one hand, domestic individuals and groups, and, on the other, a transnational order which protects the former through binding decisions against states and large-scale organisations. It also facilitates a mutual socialisation of local, national and continental jurists, refining the coherency, consistency and power of ever-more effective legal decisions.

According to Joseph Weiler, what decisively sets EU law apart from previously established international treaty law, is not the principle of supremacy itself, but its reliable and effective implementation in Europe.[23] In traditional international law, treaty provisions often have equal status with domestic law, which may be superseded by subsequent national legislation. But EU law relates

[23] See Joseph H. H. Weiler, "A Quiet Revolution: The European Court of Justice and Its Interlocutors" (1994) 26 *Comparative Political Studies* 510; and "The Transformation of Europe" (1991) 100 *Yale Law Review* 2403.

to member state law in the way that, for instance, federal law relates to the law of individual states in the USA, not as international law relates to American federal law. European law is definitively, not conditionally, supreme over member state law.

According to Weiler, the ECJ is also a better guarantor of EU policy than the European Commission, which maintains links perhaps too intimate with member state bureaucracies. On the basis of treaty provisions, the Court sets out structural doctrine on the rules governing EU and member state relations, as well as material doctrine on the economic and social aspects of that relationship. Member states seek judicial remedy, clarification on European law when necessary, and abide by it. For Weiler, these are factual, not simply normative, statements about ECJ power.

Again, individuals appeal through national courts to European law. According to Weiler, even though member state courts make the final decision on these claims, their requests for clarification by the ECJ through article 177 of the EC Treaty, increasingly Europeanises national law. National actors interact with EU law through national courts more extensively than with any previous kind of international law, making it accessible, and encouraging their stake in it. National courts do not want to violate the professional courtesy among jurists by violating or ignoring ECJ rulings. Nor do they want to seem less progressive than more enthusiastic member states. EU law has given power to the lowest national courts that higher national courts have withheld from them. In fact, EU law empowers national judiciaries as a whole *vis-à-vis* other national branches of government.

According to Weiler, national executives and legislatures do not behave antagonistically toward the ECJ because they see it as fairly neutral, they are willing to sacrifice short-term for long-term gain, and, in any case, they have so much power in making the policy that the Court decides over. Weiler claims that national courts, and to a lesser extent, legislatures and executives, accept ECJ decisions partly because they are formulated in legally-formalist terms, appearing objective and neutral. Thus the ECJ is granted latitude by the national executives and legislatures, which they previously granted to national courts under the traditional *Rechtsstaat*.

An additional reason for the surprising success of the ECJ and EU law, is that they fulfil the expectations of internationally-inclined intellectuals who were disappointed by the Cold War-era United Nations. According to Weiler, the brightness of the ECJ's future is nevertheless somewhat dimmed by an increasing backlog of cases that will cause irritation on the part of claimants, a general backlash against extensive judicial review, and increased inter-institutional EU rivalry. However, these are constraints that courts managed to overcome in the transition from the traditional to the social *Rechtsstaat*. More ominous though is the fact that majority voting in the Council of Ministers (the strongest representation of individual member state interests) means less overall consensus on policy, affecting the Court's decisions. Like national courts, the ECJ may move

more tentatively on issues where there is not clear public or institutional consensus.[24]

Anne-Marie Burley (now Slaughter) and Walter Mattli understand legal integration as a gradual penetration of EC law into the domestic law of its member states.[25] *Formal penetration* is achieved through the supranational legal acts from Treaty law to European Union law; and through cases by which individuals appeal to European law in member state courts. *Substantive penetration* is achieved through spillover from economic into social spheres (health, worker safety, welfare, education and eventually political participation). For Slaughter and Mattli, law is the functional domain that circumvents the direct clash of political interests. European actors have found it in their interest to promote incremental expansion through *functional spillover*—different economic sectors can only be well-integrated by action in spheres other than those directly involved, which in turn requires more action; and *political spillover*—economic integration encourages changes in expectations, values and strategies of national interest groups at the supranational level. Rather than an ideologically resistant obstacle to the socialisation of formal law, in the European context, the ECJ is a wholeheartedly enthusiastic, if cautious, actor in the construction of a yet still unrealised *Sozialrechtsstaat*.

Christian Joerges emphasises the normative power of European law despite its lack of lethal sanction and direct democratic legitimacy.[26] A burgeoning constitutionalisation of Europe is nothing to be surprised about, according to Joerges, as a common market necessarily equals a common constitution.[27] The ECJ speaks "legalistically" to protect itself, as Slaughter, Mattli and Weiler suggest, but, it must be understood by, and engaged with, political actors. After all, the terms of legal integration have direct impact on political reality. Borrowing a conceptual trope from Michelle Everson,[28] Joerges explains how the legal integration process is simultaneously one of political disintegration as well: for instance, member states who rely on minimal environmental standards experience pressure to conform with higher ones; member states with high social welfare levels are somewhat disadvantaged competitively in an open market. As

[24] See Gerald N. Rosenberg, *The Hollow Hope: Can Courts Bring About Social Change?* (Chicago: University of Chicago Press, 1991).

[25] Anne-Marie Burley and Walter Mattli, "Europe Before the Court: A Political Theory of Legal Integration" (1993) 47 *International Organization* 41. See also Anne-Marie Slaughter Burley, "New Directions in Legal Research on the European Community" (1993) 31 *Journal of Common Market Studies* 391.

[26] Christian Joerges, "Taking the Law Seriously: On the Political Science and the Role of Law in the Process of European Integration" (1996) 2 *European Law Journal* 105.

[27] In what seems to be a confirmation of Marxist charges that national constitutionalism was the mere product of the nationalisation of markets in the eighteenth and nineteenth centuries, Joerges assumes that regional marketising in the EU brings about regional constitutionalisation. Whether such marketising will have the devastating economic effects on vast portions of the European populace that did the first two centuries of national marketising in individual states before the establishment of the *Sozialrechtsstaat* is something to be addressed below.

[28] Michelle Everson, "Laws in Conflict: A Rationally Integrated European Insurance Market?" (Ph.D. dissertation, European University Institute, Florence, 1993).

external standards change, internal regulations collapse and crumble. Joerges assumes an ongoing repetition of the development pattern of the national market/constitutional state at the continental level. Thus he predicts the second coming of legally-enabled process of constitution-building whereby macro-level economic and political development proceeds at the expense of obstructionist local institutions.

Joerges explores the ramifications of the October 1993 Maastricht Decision of the German High Court that struck a blow to ECJ-led integration.[29] The *Bundesverfassungsgericht* declared that: (1) the EU is a state association suspended *between* a confederation and a state-like entity; (2) if the ECJ extends EU power itself, without a Treaty amendment, such judgments will not have binding force in Germany; and (3) the German Court defines "the people" as a *Staatsvolk*, a state-people, making it nearly impossible to define democracy beyond the state. In other words, the German High Court seems to roll back the traditional *Rechtsstaat* status of the ECJ and virtually rules out its development into a *Sozialrechtsstaat*. Law, as suggested by the very etymological make-up of the word, *Rechtsstaat*, cannot be substantive without a state order, which the EU, by its own profession, is not. The decision declares the ECJ to be an institution not *remotely* autonomous of other EU institutions and, by extension, the member states. Neither EU law nor the ECJ can be then understood in terms of traditional rule of law notions according to the logic of the German High Court decision. Moreover, in the invocation of the nationalist basis by which the European welfare state was largely justified, the German Court undermines the possibility of an European *Sozialstaat*, and hence an undergirding *Sozialrechtsstaat*.

While this is a potentially grave foreboding of the kinds of retrenchism faced by ECJ-driven supranational expansion of substantive rule of law in the European context in the future, Joerges remarks on the extent to which the decision did not have any immediate practical effect. The spectre of nationalist judiciaries asserting themselves against the ECJ in the future is mitigated for the ECJ-sanguine scholars by the fact that the latter has had powerful allies in local courts who may serve as a domestic constituency against higher-level appellate and constitutional courts, as well as against the legislative and executive arms.

Alec Stone (now Stone Sweet) is perhaps the most chauvinistic and unapologetic advocate of what might be called "Eurocracy through jurocracy". Consistent with other pro-Court scholars, Stone Sweet credits the ECJ with negative integration through the dismantling of the barriers to the free movement of goods, people, services and capital within Europe by encouraging states to *renounce* certain authority; and positive integration by creating new rules and institutions to regulate developments among states.[30]

[29] The Maastricht Decision of the German Constitutional Court is perhaps the definitive statement of what is called Euro-scepticism (*BVerfGE* 89, 155). Consult the responses to this decision by J.H.H. Weiler, Neil MacCormick, U.K. Preuß, and Dieter Grimm in (1995) 1 *European Law Journal* 1.

[30] Alec Stone, "Governing With Judges: The New Constitutionalism", in Jack Hayward and Edward C. Page, *Governing the New Europe* (Durham: Duke University Press, 1995), p. 286.

Stone Sweet concentrates on two facets of legal power mentioned above: the preliminary reference practice established by article 177 of the EC Treaty; and the principle of supremacy established in the *Costa* case (Case 6/64, 1964 ECR [1964] p. 585). However, further updating, expanding and documenting the Slaughter, Mattli and Weiler theses, Stone Sweet argues that the ECJ expands its power through appeals by individuals through national court references to the ECJ, and through spillover from narrow economic spheres to broader ones like gender equity, environmental protection and taxation policy.[31]

Intergovernmentalism, the thesis which maintains that the Court is ultimately the agent of the member state governments, assumes a lowest common denominator of minimal integration as a result of unanimous or super-majoritarian voting in the Council of Ministers.[32] In other words, integration is driven only so far as some of the member states less in favour of it will allow. But Stone Sweet suggests that litigation through national courts with appeals to European law *raises* this lowest common denominator by eliminating negative boundaries to integration, and addressing the lack of full compliance with European directives. He claims to demonstrate how EU law Europeanises the least Euro-friendly states from within thus changing their preferences in Council voting.

Stone Sweet asserts that member state preferences do not in fact predict Court decisions (Commission briefs, for instance, are a much better indicator); and, that even in the most integration-unfriendly context parliamentary sovereignty has been "swept aside" in areas of EU supremacy: Tory governments have asked the British Parliament regularly to amend United Kingdom statutes to conform with EU law.[33] With this invocation of intergovernmentalism and spirited attempted refutation of it by Stone Sweet, we should turn to its substance. Its arguments strike at the heart of the pro-ECJ scholars by attempting to show that ECJ autonomy and efficacy exists largely structured at the pleasure of the member states.

IV. EUROPEAN RULE OF LAW CONSTRAINED

According to game-theoretical/rational choice approaches to European law, the ECJ codifies the preferences of the major powers into law to prevent negative member state action against itself and insure compliance with its decisions.[34]

[31] Alec Stone Sweet and Thomas L. Brunell, "Constructing a Supranational Constitution: Dispute Resolution and Governance in the European Community" (1998) 92 *American Political Science Review* 63 at 72. See also Martin Shapiro and Alec Stone, "The New Constitutional Politics of Europe" (1994) 26 *Comparative Political Studies* 397; and Stone, "What is a Supranational Constitution?: An Essay in International Relations Theory" (1994) 56 *Review of Politics* 441.

[32] An articulate and very well-researched representative of this position is Andrew Moravcsik, *The Choice for Europe: Social Purpose and State Power from Messina to Maastricht* (Ithaca: Cornell University Press, 1998).

[33] Stone Sweet and Brunell, "Constructing a Supranational Constitution", n. 31 above, at 74–6.

[34] See Geoffrey Garrett, R. Daniel Keleman, and Heiner Schulz, "The European Court of Justice, National Governments, and Legal Integration in the European Union" (1998) 52 *International*

Representatives of this approach recognize the singularity of the ECJ in the international scene, finding more similarities with the United States Supreme Court than other international judicial bodies, such as the International Court of Justice, or the NAFTA or WTO dispute panels. The Court functions in a way that certainly approximates the functioning of a federal constitutional court as opposed to that of an arbitration body. But Geoffrey Garrett and his consociates, especially, argue that the ECJ is not nearly so autonomous nor capable of generating the normative structure of the EU as the ECJ-friendly scholars suggest.

Over the years, Garrett has come to refine what he considers the Court's strategic gauging of the preference positions of the member governments: where there is clear precedence for activism, the ECJ does not worry too extensively about the member states. But where the domestic cost of Court activism is high for the member states, individual states will not comply, or the concerted action of the states will react by changing the very terms by which the Court adjudicates through Treaty revision. Both discourage ECJ activism: non-compliance is embarrassing and circumvention is emasculating.

The Court is faced with the prospect of losing legitimacy in three ways: by appearing to sacrifice consistent rule of law adjudication to the wills of the member states should it always rule in obvious accord with member state positions; or through the non-compliance of individual states should the ECJ rule against the latter's will; or by a circumvention of its powers through Treaty amendment under the same circumstances. Moreover, qualified majority voting in the Council of Ministers makes it easier to undermine the Court through secondary legislation that waters down their decisions thus further deterring it from activism.

Garrett argues that the ECJ will be activist on Treaty measures where the collective action of member states is difficult (e.g., *Sheep Meat Cases*, Case 232/78, 1979). In the history of pension cases that address gender inequality, on the other hand, the Court has backed off when faced with Treaty circumvention, and formulated open-ended decisions, giving it more room to manoeuvre in subsequent cases. Open-ended decisions are one of the chief sources of the supposed legal indeterminacy that arises when courts preside over socially complicated and politically controversial cases. As the argument goes, they either fail to get the issues correct despite good faith, or adjudicate in a deliberately vague manner so as to not rouse the ire of the more powerful branches of government.[35]

Organizations 150; Garrett and Barry Weingast, "Ideas, Interests and Institutions: Constructing the European Community's Internal Market", in Judith Goldstein and Robert Keohane (eds.), *Ideas and Foreign Policy* (Ithaca: Cornell University Press, 1993), p. 193; and Garrett, "The Politics of Legal Integration in the European Union" (1995) 49 *International Organizations* 171.

[35] For recent and various formulations of these arguments, see Richard Posner, *The Problem of Jurisprudence* (Cambridge, MA: Harvard University Press, 1990); Roberto M. Unger, *What Should Legal Analysis Become?* (London: Verso, 1996); Cass R. Sunstein, *Legal Reasoning and Political Conflict* (Oxford: Oxford University Press, 1996); Duncan Kennedy, *A Critique of Adjudication: fin de siècle* (Cambridge, MA: Harvard University Press, 1997); and Rogers M. Smith, "The Inherent

The ramifications of Garrett's arguments that the ECJ seeks to avoid confrontation with other EU institutions is that the adjudication of EU law will remain exclusively confined to issues where there is member state consensus, i.e., those related to an integrated market. Moreover, in those cases that actually are heard and decided by the ECJ, adjudication will be conducted under the apprehension that the Court's decisions will remain sufficiently muddy in the hope that the more political EU institutions may do as they see fit with them. This is a clear compromise of standards of the traditional *Rechtsstaat* and, on this basis, aspirations for the consolidation of a European *Sozialrechtsstaat* may be ruled out altogether. The Court will certainly be discouraged from dealing with issues of substantive economic equality, such as disparity between pensions earned by different genders within one state or varying degrees of social insurance among member states. The priority of economic "liberty" over substantive "equality" in supranational adjudication potentially sets the rule of law back over a century in the European context. The *Sozialrechtsstaat* will have reverted to an even further scaled-back capitalist *Rechtsstaat*, since the ECJ would have even less autonomy than nineteenth century courts to adjudicate the social consequences of a free market.

V. PROSPECTS FOR AN EU RULE OF LAW

All of the students of European law treated in this essay recognise a newly-emerging juridical reality in the EU and new normative possibilities concomitant with it. From the perspective of the ECJ-sanguine scholars, the likelihood of a fully liberal or social democratic rule of law in Europe is exceptionally good. The ability of European citizens to appeal to a court above the governments of the nation-states in which they live—even if through local courts—is a potential safeguard against nation-state violations of rights, or even an encouragement of those states to expand rights. The ECJ is a supranational court with greater power to guarantee cosmopolitan norms than international courts of human rights have been, since it is more intimately tied to domestic orders.

The mutual interaction and socialisation of jurists from all levels and countries of the EU that these scholars point to as the source of an ever-more norm-refining professional process may guarantee a more coherent and effective international-federal order than any that has ever existed before. On the other hand, these lawyers might be forgiven for exaggerating the substantively progressive possibilities of expanded international discourse among jurists, and underestimating the inherent democratic deficiencies of such an outlook.

The boldest claim of the pro-ECJ scholars is that the economic sphere of adjudication explicitly granted to the Court is presently "spilling over", and will

Deceptiveness of Constitutional Discourse: A Diagnosis and Prescription", in Ian Shapiro and Robert Adams (eds.), *Nomos XL: Integrity and Conscience* (New York: New York University Press 1998), p. 218.

increasingly do so, into social and political spheres. The Court will continue subtly to argue that economic integration cannot be achieved without social spheres being integrated as well. A fully common market cannot function properly without universal levels of private and social insurance, environmental protection, medical benefits, worker compensation, etc. Thus will the EU become a supranational social democracy through law.

Therefore, the mutual socialisation of judicial actors from different levels within, and from across, European member state borders, and the institutional autonomy claimed for the ECJ by the pro-Court scholars seems to suggest that the traditional characteristics of the rule of law are preserved at the EU level: along the lines of the traditional *Rechtsstaat*, independence from other institutional actors and consistency of adjudication obtains throughout the system. The rosy picture painted of social spheres integrating through economic ones would again confirm that a European *Sozialrechtsstaat* is possible and likely. Needless to say, the Eurosceptic scholars give us reason to pause about such conclusions or predictions.

The work of the "juro-sceptics" suggest that a rule of law other than one confined exclusively to economic integration is not likely to emerge in the near future. Moreover, given the lack of autonomy attributed to the Court in this literature, serious questions are raised about the ECJ's ability to conduct even such a narrowly defined task as strict common market-making by standards of the traditional rule of law. One of the ramifications of this perspective is that European law will function like nineteenth century national law in the USA, adjudicating only the contemporary equivalent of property rights and free trade—with all the business-friendly implications that this entails. But it may do so without the consistent and good-faith laissez-faire perspective maintained by most jurists in the context of the nineteenth century, a perspective that judicial actors were able to maintain in practice by appeals to constitutional or rule of law norms and the factual condition of their institutional independence.

This lack of ECJ autonomy in the juro-sceptic presentation obviously also implies that the Court will be unable to surreptitiously socialise European and domestic law in a progressive manner. To be sure, national courts have not and do not function free of institutional constraints,[36] but because the ECJ, while more entrenched in domestic orders than other international courts, nevertheless remains somewhat remote from national institutional structures, it seems to be especially vulnerable to such pressure.

As inspiringly optimistic as are the projections of the Court-friendly literature, those that can be drawn from the ECJ-sceptical literature are bleak. If the governing principles of a juridical unit like the EU in fact merely pertain to the regulation of an economic free-trade zone, those principles will clearly be too thin to preserve the gains of liberal or social democracy associated with two or three centuries of expanding civil and social rights. This is especially so if the

[36] See Rosenberg, *The Hollow Hope*, n. 24 above.

judicial arbiter of them, the ECJ, lacks the necessary autonomy to keep the other political institutions from enacting arbitrary and inconsistent policies—perhaps one of the few attributes of the laissez-faire judiciaries of the nineteenth century worthy of admiration.

The need for supranational legal protection in response to the expanding transnational power of economic and hence political actors would be, according to the Court-sceptical model, stymied by the still quasi-sovereign nation-states. The EU member states may have suffered some loss of sovereignty to the extent that they participate in the process of European integration as an attempt to compensate for economic losses that some would attribute to international developments related to economic globalisation.[37]

However, these states may be still sufficiently strong, and exceedingly jealous of what sovereignty they actually hold, so as to be able block the extension of social principles of justice from entering the adjudication processes of the ECJ. The fact that economic issues are the only ones to be adjudicated at a supranational level in Europe would not be so problematic if member states remain capable of commanding a monopoly on the legal protection and enforcement of social rights domestically.[38]

If, however, economic control at the domestic level is abdicated to any significant extent in the future due to the pressures of globalisation, an abdication of social and political control would be expected to follow as well. This would leave no legal forum, national or Europe-wide, in which social issues are adjudicated decisively. This is the looming danger to the *Sozialrechtsstaat* at all levels of European governance should the ECJ-sceptical prognostications prove to be true.

While it is fashionable to indulge in Cassandra-like pronouncements concerning the demise of both the welfare state and the rule of law as a result of the growing transnational mobility of capital, it is nevertheless important to consider worst case scenarios: in the gap between the European member states' declining ability to secure and advance principles of the *Sozialrechtsstaat* themselves (as a result of evaporating tax bases, increased environmental threats, anti-immigrant and minority-unfriendly policies, etc.), and their reluctance to fully accede these responsibilities to fora like the ECJ, lies the abyss of the supranational democratic possibilities of the European social rule of law. If the ECJ is to adjudicate exclusively on the basis of market-related rules, and merely

[37] See Paul Hirst and G. Thompson, *Globalisation in Question: The International Economy and the Possibilities of Governance* (Cambridge: Polity Press, 1996); Ash Amin and Nigel Thrift (eds.), *Globalization, Institutions, and Regional Development in Europe* (Oxford: Oxford University Press, 1996); and Saskia Sassen, *Losing Control?: Sovereignty in an Age of Globalization* (New York: Columbia University Press, 1996).

[38] Garrett himself, for instance, does not draw such generally apocalyptic conclusions for social democracy in the EU because of his belief that globalisation is not adversely affecting, but rather, on the contrary, is enhancing, domestic social welfare policy in the European member states. See Geoffrey Garrett, "Global Markets and National Politics" (1998) 52 *International Organization* forthcoming.

hortatory appeals to principles associated with civil, social and human rights, then the subtle distinction and interplay of economic adjudication and social adjudication in the Court-friendly "spillover" model would appear to be virtually ineffectual in the EU context. On the other hand, the Court-friendly scholars have amassed serious evidence to suggest that a constitutional order, as rich as any national liberal or social democratic model, is emerging in Europe and is making possible the transposition of the *Rechtsstaat*, and eventually the *Sozialrechtsstaat*, to a supranational level.

These normative speculations depend on further empirical work like that examined in this essay. However, the scholarship on law and the Court in the EU already provides us with the framework for conceptualising the future of the rule of law in one important, if perhaps exceptional, segment of an increasingly supranationally-organised world.

14

Constructing Law's Mandate

KENNETH WINSTON

Orderliness is a demand we make on the world, William James observed. This is not to say that the world in itself is chaotic, lacking a prior order, but it may not be the order we desire or need. Our task is to undo the original order and construct one suitable to our purposes, to our human interests. But what do we count as orderly or in order? That depends on what our purposes and interests are. Thus, we cannot decide the question of order in the abstract. Lon Fuller offers a helpful illustration. Consider, he says, that a group is in order for walking a mountain trail if the members are lined up behind one another; they are in order for a military attack across a field under fire if they are deployed; and they are in order for a photograph if arranged in a more or less random-appearing, "natural" way. In each instance, the apparent judgement of fact ("the group is in order") is a tacit appraisal of the existing arrangement as a means to an end (e.g., to confront the enemy). Recognition of an arrangement as orderly turns, then, on an evaluation, revealing that certain judgements of fact are not what they appear to be.[1]

Fuller found the same kind of evaluation in the work of legal theorists when they assert that a legal order exists. Each theorist sets out criteria, often tacitly, for what makes a social activity sufficiently in order to be called (by the theorist) a legal order. Among a number of prominent theorists in the twentieth century, for example, the dominant criterion has been that law must take the form of a "system", involving a master test for determining which norms are included in the system—and hence binding on legal subjects—and which are not. Any social arrangement lacking this feature fails to qualify as a legal order. But the existence of law, as Fuller remarks, is unlike that of apples and comets. When theorists deploy criteria of orderliness, they are not simply mirroring an independent fact-of-the-matter; they are, to a degree, constructing the phenomenon. And since judgements of what is in order depend on assumed purposes, no theorist escapes making an evaluative judgement in declaring that a legal order exists. Moreover, any theorist who assumes a different end, or a different assessment of the relation between means and ends, may disagree on the apparently factual

[1] I discuss Fuller's illustration and its implications in "Is/Ought Redux: the Pragmatist Context of Lon Fuller's Conception of Law" (1988) 8/3 *Oxford Journal of Legal Studies* 329–49. For citations to William James, see especially at 336–7, 341, and 345–6.

question of whether or not a legal order exists. (In this connection, Fuller was fond of recalling Nietzsche's reminders about what we have forgotten, including the reminder that some facts—or what we take to be facts—are products of prior, and forgotten, valuations.) Of course, theorists typically disguise their appraisals as descriptions of independently existing states of affairs. To the extent that the work of a theorist is parasitic on the purposive efforts of a population sustaining legal practices by means of their own evaluations, the theorist may appear to be a reporter of antecedent facts. But just as a scrupulous lexicographer cannot record existing meanings without also clarifying them, just as a conscientious judge cannot articulate a rule without also changing the rule, so a responsible theorist cannot describe a legal practice without improving upon it.

Fuller, I believe, was inclined to see the work of theorists as parasitic in just this way. Law, he wanted to say, is the work of its everyday participants, a continuous effort to construct and sustain a common institutional framework to meet the exigencies of a shared existence, to resolve recurrent conflicts, and to make social life more satisfactory for all. He emphasised, to be sure, that the achievement is always liable to reversal or diversion into pathological forms. One of his most well known mythopoeic devices is the story of the hapless king, Rex, who with the best of intentions attempts to make laws for his subjects and fails. The failures are instructive, because each of the eight ways in which Rex bungles the job involves a violation of "the morality that makes law [i.e., legislation] possible". Only by adhering to the tenets of this morality can a legislator succeed in achieving governance according to rules. The larger lesson is that the enterprise of governance by rules depends, not on what theorists believe, but on what its participants take it to be. What matters is their beliefs, their commitments, their endorsement of the purposes it is designed to achieve.[2]

But if orderliness is a demand that derives from us, how demanding are we? What sort of demands do we make? Fuller's central point is that we make a moral demand; our striving is guided by certain ideals. After all, we do have choices, and any specific legal practice will lead us to conduct our lives one way rather than another. As he says: "A social institution makes of human life itself something that it would not otherwise have been".[3] Legal institutions are "imagined" practices, and institutional design is moral legislation. For that reason, the significance of a legal order lies in the moral direction it gives to human conduct. Thus, Fuller says that the controversy over the meaning of "law" is not about "mere images of some datum of experience, but direction posts for the application of human energies".[4] In this respect, Fuller reflects James' view that

[2] On this point, Fuller is in accord with A. D. Lindsay, *The Modern Democratic State* (New York: Oxford Univ. Press, 1962), pp. 37–8. The "morality that makes law possible" is discussed by Fuller in *The Morality of Law*, rev. edn. (New Haven: Yale Univ. Press, 1969), especially ch. 2.

[3] "Means and Ends" in Kenneth Winston (ed.), *The Principles of Social Order: Selected Essays of Lon L. Fuller* (Durham: Duke Univ. Press, 1981), p. 54.

[4] "Positivism and Fidelity to Law—A Reply to Professor Hart" (1958) 71/4 *Harvard Law Review* 632.

conscious life is not exhausted by what merely is; it always means something that is not yet, "an ideal presence which is absent in fact".[5] The order of law must be understood accordingly.

At the same time, because choice is possible and people may disagree on ends (and the relation between means and ends), the ideal for a specific practice is often contested. The kind of order postulated or desired may be in dispute. The fate of ideals in institutions, as Philip Selznick says, "is always problematic, never to be taken for granted".[6] So in any given instance we need to ask, whose construction are we dealing with? Do we all make the same demand? If not, whose demands count? Since the question of order cannot be treated in the abstract, a response to these questions requires close investigation of what happens at the most basic levels of an organisation. My aim here is to take some steps in this direction, through a limited examination of a specific government agency whose governing ideal has often been contested, especially in recent years. That agency is the Solicitor General's office in the United States Department of Justice. At issue is the construction of the Solicitor General's mandate.

Two views of the Solicitor General's office

I concentrate on the Solicitor General (SG) because, in connection with another project, I have been trying to assess the mandate of this office and how it can be contested.[7] Specifically, during the Reagan presidency, the Department of Justice exercised its formal authority over the SG in a determined attempt to transform the office into a promoter of the president's policies before the Supreme Court. These efforts were, I believe, largely unsuccessful, in great part because the two lawyers who served as SG during the Reagan years, Rex Lee and Charles Fried, recognised the seriousness of the threat and combatted it fairly effectively. But the story of this controversy is instructive for understanding the construction of law's order.

Of the two SGs, Charles Fried is an obvious focus of attention. After leaving office, he wrote an extended account—or apology, we might say—of his tenure as SG and brought to bear his skill as a legal theorist to offer a picture of the SG's office and of his dealings with other members of the Reagan administration.[8] As we shall see, Fried regarded his skill in legal philosophy as a crucial ingredient of his competence as SG and a principal reason for his appointment. Whether or not that was so, the virtue of Fried's after-the-fact account is that legal theory

[5] William James, *A Pluralistic Universe* (New York: Longmans, Green, 1909), p. 284.

[6] Selznick, review of Fuller's *Anatomy of the Law*, (1970) 83/6 *Harvard Law Review* 1478.

[7] Winston, "Moral Opportunism: A Case Study" in Ian Shapiro and Robert Adams (eds.), *Integrity and Conscience* (New York: New York Univ. Press, 1998), pp. 154–84.

[8] Charles Fried, *Order and Law: Arguing the Reagan Revolution—A Firsthand Account* (New York: Simon & Schuster, 1991).

enters, not systematically or for its own sake, but to help make intelligible a particular way of engaging in the ongoing enterprise that constituted the SG's office. Too often, theorising about legal institutions is done abstractly, as a kind of absentee intellectual management (to use one of Fuller's favourite phrases), deciding questions wholesale about what legal practices ought to be or do. Fried's account, to the contrary, is the work of a reflective practitioner; it introduces only such theoretical pronouncements as actually function in a practical way, to guide descriptions, actions, or evaluations. If the resulting picture of the SG's office is less sharply delineated than a legal theorist might like, it has the warrant of portraying the office as an enterprise that matters in quite specific organisational and normative ways. That is as it should be, I believe, especially since Fried himself was torn between a conception of the SG's office inherited from previous administrations and what the Reagan team was attempting to make of it. I shall suggest that Fried was right to be torn—or, rather, wrong to give as much credence as he did to his superiors' views. Although my assessment is motivated by more general concerns about legal order, I hope I shall not fail to do justice to Fried's serious engagement with the ongoing contest over the construction of the SG's office.

To set the scene for this contest, the obvious starting point is a description of the SG's role, especially as few people are aware of its functions. But immediately we confront the problem that the correct description is itself contested. Since descriptions of practices are assessments of existing arrangements in their capacity to serve assumed ends, everything depends on the ends assumed and their often-complex relations to means. To focus the discussion, I shall explore two descriptions of the SG's office, which I shall refer to as *positivist* and *naturalist*, explaining my use of these terms as I go along.[9] The central difference between them turns on the relation between law and politics—which means, in practical terms, the relation between the Department of Justice and the Presidency.

Positivists and naturalists would agree that the principal statutory charge to the SG is to conduct appellate litigation for the federal government. From there, the positivist description highlights formal lines of authority. The SG is a political appointee who serves at the pleasure of the President and who works for the Attorney General, as the latter's legal expert. Although the SG is the only government official required by statute to be "learned in the law", everything the SG does is subject to the general supervision and direction of the Attorney General, who can decide which cases to argue and which positions to adopt in those cases. Indeed, the Attorney General retains the legal authority to displace the SG and assume the role at any time.[10] In a word, since for the positivist law

[9] I first drew this distinction in "Moral Opportunism", n. 7 above, at 156–8. The present essay follows that account but elaborates the distinction and connects it to a general dispute in legal theory. For the term "legal naturalism", see Selznick, n. 6 above, p. 1475.

[10] Act of 22 June 1870, ch. 150, s. 2, 16 Stat. 162. The most recent codification of the law is at 28 U.S.C., s. 518 (1988). Joshua Schwartz points out that a similar requirement to be "learned in the

emanates from legitimate centres of power, the positivist description emphasises that the SG is a subordinate officer within the Department of Justice.

The naturalist has a different view of law and authority. For the naturalist, some legal standards are more authoritative than others, and their authority is not determined wholly by formal criteria.[11] Formal authority can become attenuated, for example, if citizens regard the content of specific laws as morally objectionable. Drug policy in the USA illustrates the point, insofar as it fails to distinguish trade in "soft" drugs for personal (and medicinal) use from trade in "hard" drugs. The result is a widespread public sentiment that many drug laws are illegitimate and need not be obeyed. Similarly, the willingness of American citizens to comply voluntarily with tax laws—which is very high—is a function of their sense of the fairness of the tax code.[12] Formal authority can also become attenuated by the way legal standards are administered, if they are selectively enforced by the police, say, or applied judicially on the basis of poor reasoning.[13] Likewise, attenuation accompanies any gap between what laws require and the circumstances in which they are meant to be obeyed. Fuller observed that respect for traffic laws depends, in part, on the way streets are laid out. In sections of Boston or Cambridge, where the maze of streets exhibits no evident principle of design, both observance of the law (by motorists) and supervision of traffic (by police officers) are undermined. Highway engineers, we could say, are architects—or wreckers—of law's authority.[14] Yet another kind of attenuation, more relevant here, occurs when the formal rules of an agency are ill-suited to the agency's general aims, producing, in the worst case, a disjunction between officials' responsibilities and the forms by which they are held accountable. In such circumstances, alternative practices more in accord with the agency's mission may develop informally, crystallise into firm expectations, and become authoritative—competing with, if not displacing, the formal rules. This scenario captures the history of the SG's office.

Accordingly, a naturalist account of the SG's role is practice-based. It focuses on operative norms, tacit assumptions, and settled expectations, especially where those diverge from the formal rules. In particular, the naturalist highlights the degree to which the separation of law and politics has acquired normative force. Let me emphasise that this separation, in the sense I intend, does

law" was placed on the Attorney General by the Judiciary Act of 1789, ch. 20, s. 35, 1 Stat. 93, but is not present in the current statute, at 28 U.S.C., s. 503 (1988). He observes: "Whether intentionally or otherwise, the difference reflects contrasting expectations that have developed concerning the degree of political insulation or involvement of these two officers": Schwartz, review of Fried's *Order and Law*, (1992) 60/4 *George Washington Law Review* 1083 n. 3.

[11] The reader should note that my aim, in this essay, is not to elaborate a theory of law's authority, only to point out that the reasons officials have for enacting and enforcing laws—or, alternatively, the reasons citizens have for compliance with official pronouncements—are not limited to content-independent reasons.

[12] Tom R. Tyler offers an empirically based analysis of obedience and legitimacy in *Why People Obey the Law* (New Haven: Yale Univ. Press, 1990).

[13] Selznick, n. 6 above, p. 1477.

[14] Fuller made the observation in a private letter to James Q. Wilson, 30 January 1973.

not mean that legal reasoning is anything other than a species of moral reasoning (I shall say more in a moment about which species it is). The important point here is that the orderly development of law occurs in distinct judicial or quasi-judicial contexts, each with its own set of constraints, which mark it off from the pursuit of a political party's agenda. The first significant step toward separation occurs in the Attorney General's office itself. Although the Attorney General is often a close adviser to the President and is not necessarily expected to exercise undistorted legal judgement, the Attorney General is responsible for safeguarding the legal mission of the Department of Justice, by protecting it on appropriate occasions from political intrusion. (Different occupants of the office, of course, have been more or less successful in executing this responsibility.) The SG is still further removed from politics than the Attorney General, and the control of professional norms is appropriately stronger. The task of the office, remember, is to conduct appellate litigation for the government. Being "learned in the law", the SG is expected to exercise that responsibility with dispassionate legal judgement. So, in practice, as the *normative history* of the office has played itself out, and as the best SGs have provided models of dispassionate judgement, the expectation has developed that the Attorney General will generally defer to the SG's opinion in specific cases. This expectation has been reinforced at the next level of the hierarchy by the SG's deference, in turn, to the small cadre of high-quality lawyers who work in the office, where the norms of professionalism are at their strongest.

What accounts for these norms of deference? The short answer is: the rule of law. Just as the rule of law requires a judiciary independent of politics, so the SG's role, on the naturalist view, requires at least some degree of institutional autonomy. Such autonomy permits the virtues of the lawyerly craft to be exercised—reasoned elaboration of legal standards over time, consistency across cases and areas of the law, fact-sensitivity, and so on. In this way, rule of law values provide the measure of law's "orderly" development.[15] By contrast, in politics (in the conventional view) anything goes. No particular form of reasoning is needed beyond the exercise of power itself or the vagaries of political persuasion. Of course, the naturalist has to avoid exaggerating the SG's independence from politics—sometimes done by former SGs themselves. Francis Biddle, who was SG from January 1940 to September 1941, has offered perhaps the most striking example. In his memoirs, Biddle wrote that the SG "is responsible neither to the man who appointed him [the President] nor to his immediate superior in the hierarchy of administration [the Attorney General]. The total responsibility is his, and his guide is only the ethic of his law profession framed in the ambience of his experience and judgment". Continuing these high-minded sentiments, Biddle said: "Nor are there any of the drawbacks that

[15] Lincoln Caplan suggests that the lawyerly virtues identified in the text have a special, not to say peculiar, connection to the "legal process" school of jurisprudence. But his reasons for this claim are unclear, and his description of the legal process school is a simple textbook version: Caplan, *The Tenth Justice: The Solicitor General and the Rule of Law* (New York: A. A. Knopf, 1987), pp. 74–5.

usually go with public work[:] no political compromises, no shifts and substitutes, no cunning deviations, no considerations of expediency. The Solicitor General has no master to serve except his country".[16] (No heavy lifting, either, one is tempted to add.) The reality is more complicated. In many cases that reach the SG's office, the best judgement of the career lawyers is that the decision could go either way as a matter of law. This should not be surprising, since many cases that make their way to the Supreme Court are likely to be close calls. For these cases, the question legitimately turns on a matter of political morality, and it is entirely appropriate for the Attorney General and the President to be involved. In these cases, one could say that the SG's obligation is to make certain of their participation. The nature of that participation, however, is another matter; I shall say more below about how political morality properly enters.

It does not follow that the SG always achieves the right balance in practice. To take one illustration: when *Reynolds v Sims* was presented to the Court for decision in 1963, a staff lawyer argued to SG Archibald Cox that the heart of the case—whether to impose on the states the federal principle of "one person, one vote"—was a question of political morality. Accordingly, he recommended that the position taken in the government's brief should be decided "at the highest levels". Cox disagreed. He conceded that "one person, one vote" was sound public policy, and he knew that Attorney General Robert Kennedy supported that position. But Cox believed that it did not have a firm basis in law. This assessment, actually, had less to do with decisions in previous voting cases than with a particular view of the Court's constitutional role. In a memo to Kennedy, Cox wrote: "In my opinion any such decree [imposing "one person, one vote" on the states] would be too revolutionary to be a proper exercise of the judicial function and too rigid to comport with the principles of federalism". And so Cox argued before the Court, without any interference from Kennedy.[17] Later, Cox acknowledged that he had misjudged the implications of the decision. In thinking it would precipitate a constitutional crisis and cause an enormous drop in public support for the Court, he had been, in his own words, a "poor prophet".[18] Cox's mistake, however, did not consist in failing to predict the Court's decision. The duty of the SG's office is not to limit itself to opinions it thinks the Supreme Court is likely to adopt. (That could be, I suppose, a legal realist view of the office, applying Holmes' dictum that the law is nothing but a prediction of what the Court will do in fact.) Rather, the duty is to provide the Court with the best reading of the law's orderly development. Cox did not see a plausible reading in the direction of the *Sims* decision, given the kind of power the Court would have to exercise—despite an alternative reading that happened to be favoured by his political superiors.[19]

[16] Francis Biddle, *In Brief Authority* (Garden City: Doubleday & Co., 1962), pp. 97–8.

[17] See Victor Navasky, *Kennedy Justice* (New York: Atheneum, 1971), pp. 309ff.

[18] Caplan, n. 15 above, p. 194.

[19] For Cox's account of his role in the reapportionment cases, see *The Court and the Constitution* (Boston: Houghton Mifflin Co., 1987), pp. 288–304.

In a naturalist's description, then, the independence of the SG consists in maintaining an often delicate balance between a fiduciary duty to the law and responsiveness to a particular administration. The SG's office is respected for its independence, not when it has no master, but when it places its commitment to one master *on a par* with its commitment to the other.[20] Total independence would actually mean a lack of democratic accountability and would expose the office to political pressures. Paradoxically, it is the Attorney General's formal authority to overrule the SG that protects the SG in exercising the fiduciary duty to law. Charles Fried understood this point and, by his own account, exploited it effectively. He reports that whenever any of his colleagues in the Department of Justice were dissatisfied with one of his legal determinations, he invited them to appeal to the Attorney General. That freed him "to take whatever position seemed right to me, while giving people a sense that a route of correction was open . . . If the Attorney General did not approve [my decision], he could overrule it".[21] So, even in ideal terms (on the naturalist account), the SG should be answerable to the Attorney General, and thus the President, at the same time that, as a matter of practice, the Attorney General should generally defer to the SG's legal judgement.[22]

A brief excursus on positivism

I am suggesting that positivists object to the naturalist account—with its accent on the SG's fiduciary duty to law and the consequent norms of deference to professional judgement—in large part because it diverges from an exclusive (or predominant) focus on formal lines of authority. An acceptable theory of law, I want to say, has to recognise rather that formal law has a variable relation to authority. Yet positivist accounts of legal institutions run in the opposite direction. They are characteristically rule-centered and feature explicit, preferably

[20] For this formulation, see Philip Heymann, *The Politics of Public Management* (New Haven: Yale Univ. Press, 1987), p. 72. Illuminating comparisons could be made to other executive branch agencies. In contrast to the SG's office would be the independent regulatory agencies, such as the Federal Reserve, the Federal Trade Commission, and the Securities and Exchange Commission, whose members, although appointed by the President, do not serve at the President's pleasure. Much closer to the SG would be the U.S. Surgeon General, who, as head of the public health service, has a statutory mandate that can conflict with the President's political agenda, as C. Everett Koop so famously demonstrated.

[21] Fried, n. 8 above, p. 199.

[22] The ideal of balance is evident in Erwin Griswold's account of his tenure as SG from October 1967 to June 1973, in *Ould Fields, New Corne: The Personal Memoirs of A Twentieth Century Lawyer* (St. Paul: West Publishing, 1992). For example, Griswold says: "[T]he fact that [the SG] is in the Executive Branch and subordinate to the Attorney General and the President is not, to my mind, sufficient to define his role. His role is basically professional. He is an 'officer of the Court' in a very real sense" (p. 326). Then he adds: "I tried to be aware of the policies of the Attorney General and the White House, and to support them within wide limits . . . From time to time I adjusted my views, and, more frequently, my phraseology, to meet points made by my superiors" (p. 329). "It can on occasion be a tightrope, but the Solicitor General should keep his balance" (p. 330).

written directives issued by official sources. They emphasise the chain of command. Why?

Philip Selznick has observed: "The most striking feature of the ethos of positivism is *a quest for determinacy*". Legal positivists abhor what is unsettled and open-ended; they prefer determinate rules and clear judgements, authoritatively promulgated. Even in their interpretation of the rule of law, they stress the regular and predictable application of rules to cases, ensuring reliability and security in decision-making. The central value, for them, is letting citizens know ahead of time how they will (or will not) be held to account, so that they may plan their lives accordingly. Thus, legal certainty is the preoccupation. In this light, "[t]hey resist the idea of 'incipient' or 'emergent' law or any blurring of the boundaries between law and custom, law and social practice, law and morality".[23] Why should this be so? One explanation might be this: positivists assume, quite rightly, that law exists only if it is efficacious. But it appears that they cannot conceive of law as efficacious unless a firm distinction is maintained between rules binding on citizens because they are legally valid (issuing from official sources) and rules that have a more ambiguous status, as customary or tacit or moral. For one thing, the latter rules are not entirely, perhaps not even mostly, under the control of those individuals who have the task of managing society's normative order.

In other writings, I have suggested that the principal reason the classical legal theorists took up their positivist pens was to strike a blow for social control. Austin is the most transparent, if most extreme, example. His preoccupation with the threat of social disorder has often been remarked upon. Added to this fear was his conviction that common moral opinion is so fractured, so full of partiality and prejudice, that people cannot be trusted to act decently. Indeed he says explicitly that what will always make law necessary and "highly expedient" is the "uncertainty, scantiness, and imperfection" of the moral rules people generally espouse. "Hence the necessity for a common *governing* (or common *guiding*) head to whom the community may in *concert* defer."[24] For Austin, in other words, the problem of social order is a managerial problem; without top-down control, social behaviour is likely to get badly out of hand. Common directives that are readily discernible and effective (because backed by sanctions) provide the "steadiness, constancy, [and] uniformity" society needs.

Austin, as I said, is an extreme case, but in intellectual matters the extreme case often illuminates an entire tradition of thought. My claim—asserted here, not established—is that, because of a preoccupation with social control, classical positivists proposed to identify law independently of any reference to

[23] Selznick, *The Moral Commonwealth: Social Theory and the Promise of Community* (Berkeley: Univ. of California, 1992), p. 50. Fuller, too, treats legal positivism as an ethos, rather than as a set of doctrinal claims. The same approach is evident in Ronald Dworkin's work, for example, *Taking Rights Seriously* (Cambridge, MA: Harvard Univ. Press, 1977), pp. 347–9.

[24] I discuss Austin's views in "Three Models for the Study of Law" in Willem Witterveen and Wibren van der Burg (eds.), *Rediscovering Fuller* (Amsterdam: Univ. of Amsterdam Press, forthcoming).

non-official (and only partially manageable) norms, including those of morality or justice. They thought they could do this by representing law as public order *simpliciter*. But since orderliness is a function of purpose, they always intended a certain kind of order. As Fuller said, it was not the order of a morgue or cemetery they were interested in; they meant an order that was good according to some assumed standard. "Law, considered [by positivists] merely as order, contains, then, its own implicit morality."[25] To be sure, this is a minimal ideal of order—in the sense of lacking any reference to justice or liberty—but not so minimal in its penchant for social control. (Everyone can site examples of officials who see the "threat of anarchy" where others see the orderliness of a democratic polity. Justice Antonin Scalia, for example, believes that were the state of Oregon to accommodate the sacramental use of peyote by Native American tribes within its jurisdiction it "would be courting anarchy".)[26]

Contemporary positivists reveal in various ways that they share the same preoccupation as the classical theorists, even if not as forthrightly—or for quite the same reasons—as Austin. H. L. A. Hart, for example, claims that the authority or normative force of law is content-independent. The reason for this claim, it seems, is that, since citizens may disagree about what the content of law should be, they need directives whose authority does not depend on their content. Provided that a law is issued in accordance with legitimate law-making procedures, it is binding. (This entails that the law is binding even when its purpose provides citizens with a reason to violate it. I shall return to that point in a moment.) Again, Hart asserts that a distinguishing feature of law is "the general claim it makes to priority over other standards".[27] This is another way of asserting content-independence, reflecting the positivist impulse to replace the uncertainties of moral rules or tacit expectations with known, officially managed standards. Applied to the SG's office, these claims entail that the statutory provisions establishing the office are authoritative regardless of what they provide—that, for example, the subordination of the SG to the Attorney General takes priority over the norms of deference even if the latter better serve the mission of the office.

What is most striking about Hart's formulation is the way it reifies the law; he says that "the law" makes a claim to priority. Even theorists who distance themselves from positivism as a set of doctrinal claims may exhibit the positivist ethos, particularly the tendency to reification. They, too, fail to see the authority of law as a variable attribute. In an analysis of law's authority, Philip Soper defends the claim of content-independence and talks about "the law's view" or "the viewpoint of the [legal] system itself".[28] When his analysis becomes a bit

[25] "Positivism and Fidelity", n. 4 above, at 645.

[26] *Employment Division, Department of Human Resources of Oregon v Smith* 110 S. Ct. 1595, 1605 (1990).

[27] H. L. A. Hart, *The Concept of Law*, 2nd edn. (Oxford: Clarendon Press, 1994), p. 249.

[28] "Legal Theory and the Claim of Authority" (1989) 18/3 *Philosophy and Public Affairs* especially pp. 220 and 228. For a more complete, but earlier exposition of his views, see *A Theory of Law* (Cambridge, MA: Harvard Univ. Press, 1984).

more concrete, however, "the law's view" becomes intermixed with the views of individuals. Without referring to any actual legal cases, but by way of elaboration, Soper asks us to imagine motorists better able than the legislature to assess the risk posed by an instance of speeding and, on that basis, challenging the law as applied to them. "Could a society make room for such challenges and still maintain a rule-based system of law?" Soper thinks the answer is evident. Such challenges to the law cannot be admitted, "not because it is assumed that the law has made the best balance of the underlying reasons, but because the law is binding whether or not it is correct".[29] (Is "whether or not it is correct" the same as "no matter what it says"? I think these are not easily distinguished.) Presumably, it is the *legislators* who have not made the best balance of reasons, that is, they have enacted a law that may fail to meet their own intended aims. Why does Soper think their promulgation is binding nonetheless? My hypothesis is that, despite the repudiation of Austin, Austin's ghost lingers in the form of regarding the principal function of law to be social control—and therefore laws are top-down directives. Thus, Soper says that authoritative laws are "like commands" and require action even if the source of the law is mistaken in its evaluation of the action.[30] In the ordinary sense of the word, the recipient of a command forms an intention to perform the commanded act only after the command has been issued—and would otherwise have acted differently. To revert again to Austin's view, if legal subjects are left to their own inclinations, they will engage in all manner of disorderly behaviour. Hence, social order requires stable external direction. Interestingly, however, in the one context where commands are thought to have their archetypal use, namely, the military, the exercise of leadership may depend as much on the ability to persuade as on the ability to command.[31]

Soper's motorist example implies that the content of law is arbitrary. What matters is only that everyone obey the same set of rules, in a situation where rules are needed, whatever the rules may be. The function of law is pure coordination, and rules of the road, it is thought, are paradigmatic. But some rules of the road fail to fit this analysis—and bring into view an alternative source of authority. Consider a well-known New York case, *Tedla v Ellman*, which involved pedestrians who walked on the right-hand side of a highway lacking sidewalks, contrary to an ordinance requiring them to walk on the left side, facing the traffic. They did so because the traffic on the left side was very heavy at the time, whereas the traffic on the right side was quite light. Thus, although walking on the "wrong" side, they were, in the court's words, exercising "such care for [their] safety as a reasonably prudent person would use" and as the ordinance was designed to foster. Despite this care, the content-independence thesis requires us to say that the ordinance was binding on the pedestrians. The Court

[29] Soper, "Legal Theory", n. 28 above, at 229–30.

[30] Ibid., at 218.

[31] See N. Fotion and G. Elfstrom, *Military Ethics: Guidelines for Peace and War* (Boston: Routledge & Kegan Paul, 1986), pp. 58–62.

actually held otherwise. It said, if the ordinance were binding, the legislature would have decreed that pedestrians must observe a general rule prescribed for their safety, even under circumstances where observance would subject them to imminent danger. It is unreasonable, the court thought, to ascribe to the legislature such an intention.[32]

The *Tedla* Court is suggesting that to follow a rule faithfully—what Fuller referred to as "fidelity to rule"—means reading its requirements in light of its rationale. In other words, the rationale (hence, the content) figures into determining what one is supposed to do. But even if rule and rationale are distinct, the latter has normative force as well as the former. In a particular case, therefore, the balance of reasons could go against following the rule—which means that content is relevant to the authority of a law.[33] More generally, the Court affirmed the legally cognisable power of citizens to judge the applicability of rules to their own cases. It did this by appealing to a pervasive, if tacit, understanding about rules, namely, that nothing in their nature requires that they be applied rigidly, regardless of circumstances. Whether rigid application (or, judicially speaking, "strict construction") is desirable is a policy question. In many areas of criminal law, for example, rigid application is desirable because important values such as liberty and protection from violence are at stake. But that consideration does not apply to rules of the road. Such rules should be construed as applying only to ordinary circumstances, and are properly superseded when the purpose of the rule would be better served by a violation. The Court recognised, in other words, that it was dealing with the kind of circumstance in which citizens may challenge the existing rules in just the way that Soper found inconceivable.[34] Such a challenge, of course, presupposes that citizens have the requisite capacity to apply rules, or not apply them, with intelligence. In rendering its decision, the Court expressed its confidence in citizens' capacity for rule-guided conduct.

Some theorists hold that deference to a law is more compelling—and the law's authority greater—when citizens are more likely to make mistakes in exercising their reasonable judgement in following it. I am suggesting, rather, that a law's authority is weakened when its application fails to accommodate reasonable citizen judgement, especially when based on the law's purpose. As Fuller said, a law's authority is the product of an interplay between citizens and gov-

[32] *Tedla v Ellman* 280 N.Y. 124, 19 N.E.2d 987 (1939), reprinted in Philip Davis (ed.), *Moral Duty and Legal Responsibility* (New York: Appleton-Century-Crofts, 1966), pp. 96–101.

[33] How could one think otherwise? Gerald Postema affirms "the important point", which he attributes to Austin, that "we can object to laws on moral grounds without challenging their . . . authority as laws". I think Postema is able to say this only by conflating the validity of a law and its authority. But the fact that a law was promulgated in accordance with accepted procedures is only the formal aspect of its authority. At the same time, when the content of a law is morally objectionable, officials may still have moral reasons to enforce it, and citizens may have moral reasons to comply. See Postema, "Jurisprudence As Practical Philosophy" (1998) 4/3 *Legal Theory* 352.

[34] Nor does the challenge stop necessarily with an appeal to the rationale. For, the rationale of a specific law may not accord well with other laws regulating the same sphere of activity; so the rationale may also be challenged, and so on.

ernment, not a one-way projection imposing itself from above.[35] That idea is a stumbling block for theorists who place more confidence in "the law"—and the officials who manage it—than they do in citizens. (Similarly, their worries about official discretion lead them to place more confidence in higher than in lower officials, in legislators more than in judges.) I am not denying, needless to say, that the *Tedla* Court exercised its power to review a citizen's judgement of what the ordinance required. Nor am I denying that the law is a common point of reference—or standard of rectitude—in a world where each person might prefer to follow an idiosyncratic code of conduct. If Anna Tedla had not been able to make a reasonable argument on the basis of the law's purpose, her appeal would have failed. But she succeeded, I want to suggest, because the Court could affirm the proposition that content makes a difference for a law's authority. That supports the naturalist's account of law's mandate.

Charles Fried's philosophy

The implicit morality of the positivist description of the SG helps to explain its appeal. It reflects a moral concern with stability, hierarchy, and rule following. Interestingly, however, there is a curious twist in this instance. Although positivists generally prefer a picture of law as an autonomous system of rules (autonomous from morality, custom, settled expectations—and politics), the focus on formal authority in the SG's office collapses the distinction between law and politics. Indeed it is just this separation that the Reagan administration contested. For example, in 1983, to reinforce his control, the Attorney General created a new position within the SG's office: Counselor to the Solicitor General.[36] The restructuring occurred when Rex Lee recused himself from a case and delegated to a career lawyer the decision as to what position the government should adopt. The Reagan team felt that, when the SG removes himself, he should be replaced by a "trusted lieutenant" of the President, rather than a staff lawyer whose reading of the law may not jibe with the administration's. (The new Counselor became known as the SG's "political deputy"—or, one could say, political enforcer.) This reasoning reflected the administration's political conception of the role.[37]

Even more tellingly, the administration decided to monitor the SG closely, by having Assistant Attorney General William Bradford Reynolds, the most trusted of trusted lieutenants, sit in regularly on discussions in the SG's office. The aim was to ensure that briefs in politically sensitive cases reflected the

[35] See *The Morality of Law*, n. 2 above, p. 204.

[36] Caplan, n. 15 above, p. 62.

[37] Rex Lee's recusal came in *Bob Jones University v U.S.*, 461 U.S. 574 (1983). At the bidding of administration ideologues, the Attorney General set aside the proposed brief written by Deputy SG Laurence Wallace in favour of a brief supporting tax exempt status for Bob Jones University as a religious school, despite its refusal to admit black students. With great public embarrassment, and only Justice Rehnquist dissenting, the Court rejected the administration's position.

administration's point of view.[38] A case could be "politically sensitive", of course, without being at all legally problematic. The administration wanted government briefs to be more like position papers, putting the administration on record publicly about matters of political importance, rather than presenting arguments about decisional law—which might not support the position advocated. Instead of being a counselor to the Court exercising the fiduciary duty of the office, Reynolds' efforts actually moved the SG closer to being an antagonist of the Court.

Charles Fried sometimes engaged in such political machinations, and it is to his credit that he admits doing so. (The exceptions, I believe, do not change the overall picture of Fried's resistance to his bosses.) When Rex Lee resigned in 1985, after much criticism from conservatives for his independence, Fried became acting SG. During this period, he took steps to assure the administration of his reliability. In particular, he agreed to submit a brief in an abortion case asking the Supreme Court to overrule *Roe v Wade*, even though that landmark decision was not at issue in the case. This good deed, according to Fried, placed him "in a leadership position" within the Department of Justice and secured his regular appointment as SG. In the retrospective account in his book, Fried offers an unpersuasive argument as to why *Roe* was wrongly decided and ought to have been overturned, but he admits that the decision was not one that especially disturbed him. To the contrary, it was a libertarian decision in accord with his own philosophy, and he expresses regret that the Reagan administration gave so much attention to the abortion issue.[39] This is indeed one of several examples in which Fried reveals that he did not share Reagan's moralistic streak— his willingness to employ the heavy hand of government to impose the moral beliefs of a segment of the population on the rest. (The other prominent example is Fried's criticism of the Court's decision in *Bowers v Hardwick*, which permitted states to criminalize homosexual acts between consenting adults. Fried calls Justice White's opinion for the five to four majority "stunningly harsh".) Nonetheless, at his confirmation hearing, Fried went a considerable way toward affirming the positivist conception of the SG's office, emphasising the formal lines of authority. The background to this performance was Rex Lee's fairly public, if infrequent, declarations about maintaining his independence. Lee was explicit about the threat as he saw it. On one occasion, he said: "One of the most important jobs I have [as SG] is protecting the tradition of John W. Davis, Robert H. Jackson, Charles Fahy, and Thurgood Marshall".[40] This list of names, I would note, includes some of the most distinguished SGs in American history. By including both Davis and Marshall, who were the principal lawyers on the opposing sides in *Brown v Board of Education*, Lee's statement emphasises the SG's political independence.

[38] Caplan, n. 15 above, p. 154.
[39] Fried, n. 8 above, pp. 75–88.
[40] Caplan, n. 15 above, p. 76.

Fried elaborated a different idea of independence—one of judgement, not authority—when testifying before the Senate Judiciary Committee. He said that the SG serves as an adviser to the Attorney General, "giving his own best independent judgement [about the law]", but the Attorney General, he stressed, need not accept that judgement. In the event of disagreement—which he hoped would be rare—the Attorney General "has the clear statutory authority to direct the Solicitor General to take a contrary position". When that happens, "I think it would be peevish and inappropriate for the Solicitor General to be anything but cheerful in accepting the reversal".[41] Thus, publicly Fried presented himself from the beginning (and even after he left office) as a loyal member of the Reagan team, committed to its political agenda—and its conception of the SG's role. And it was during his tenure that the Attorney General's office, under the leadership of Edwin Meese, was at its most intrusive. Given these circumstances, the common journalistic judgement, that Fried was "widely regarded as having been too much a political lackey of the Reagan Administration", is not surprising.[42] The most scathing criticism came from Lincoln Caplan, who described, among other things, a number of stratagems Fried employed that appear to have denigrated the views of the career lawyers in favour of his political superiors. Instead of recognising the special competence of his staff as "learned in the law", he regarded them, Caplan suggests, as one voice among many in the mix that determined the government's position.[43] Actually, I suspect that Fried's machinations can mostly be explained as efforts to preserve the professional norms of the office in the face of an administration extremely hostile to its mission. (Caplan makes no attempt to conceal his opposition, on political grounds, to the Reagan agenda, but he seems to confuse Fried's strategic posturing with faulty legal judgement.) In my view, Fried is to be commended for his success, such as it was, at keeping the Attorney General at bay. Though publicly he professed a commitment to the positivist conception, in practice he may have followed the naturalist conception as much as he was able. The reason there is doubt about his performance is that he raised considerable suspicion by his own public commentary, first at his confirmation hearing and then, most importantly, in the apology written after he left office—"surely", in the words of Erwin Griswold, "the most 'political' view of the Solicitor General's role that has yet appeared in print".[44]

At his confirmation hearing, Fried said: "I suppose that the President has nominated me because he has some sense of what my philosophy is, and that

[41] Fried, n. 8 above, p. 244 n. 46. Joshua Schwartz, n. 10 above, observes: "The Solicitor General [in Fried's view] is free to *propose* the positions that seem proper to him, knowing that the Attorney General is free to overrule those positions, and that the President is free to remove the Solicitor General if those positions are not acceptable. Missing from Fried's book is any set of principles to constrain the exercise of such authority" (p. 1113).

[42] Michael Winerip, "Ken Starr Would Not Be Denied" *The New York Times Magazine*, 6 September 1998, p. 51.

[43] Caplan, n. 15 above, p. 225.

[44] Griswold, n. 22 above, p. 331.

philosophy enters my judgment of what the law is on a particular matter, and what arguments should be made".[45] Now, this statement is remarkable for two reasons. The first relates to the content of the philosophy by which Fried proposes to judge "what the law is". For, it is a philosophy that put him at odds with much of the previous thirty to forty years of American constitutional development. He was alarmed not just by "ill-conceived [judicial] decisions" of the post-New Deal era but also by, in his words, "really silly statutes" which were passed, by both political parties, during the same period.[46] Fried's target here is the "liberal regulatory agenda", which involved "the aggrandizement of government" in all its branches since the New Deal, an explosive and unwarranted expansion of government's responsibilities in which the courts were complicit.[47] With the "Reagan revolution", however, this legal history would be brought to an end. That meant not just reversals in this or that landmark case, but a wholesale reversal—or at least an attempt at one—in the very conception of democratic government, from a regulatory to a libertarian orientation. And precisely because a change in legal philosophy was at the heart of the Reagan revolution, Reagan needed a legal philosopher as SG. As the official most "learned in the law", Fried would be at the centre of the revolution. He was appointed, he observes, because of his skill at doing legal philosophy.[48]

With regard to the Court's complicity in the post-New Deal era, Fried's complaint is not about judicial activism (that is, judicial decisions overturning settled law)—so long as it is for the sake of cutting back on the liberal regulatory agenda. Sometimes he presents his view differently. When describing the components of the Reagan revolution, the first tenet he enumerates is: "[C]ourts should be more disciplined, less adventurous and political in interpreting the law".[49] But he quickly makes it clear that by being political he has in mind promoting government aggrandisement. Apparently, a commitment to scaling back government is not political. Turning to the SG's office, Fried distinguishes between political commitments (which, he says, are impermissible for the SG) and philosophical commitments (which are permissible). Not surprisingly, the distinction is difficult to maintain. He himself refers to the prevailing regulatory philosophy as a "political bias".[50] Presumably, the same could be said of its opposite. However, the error in Fried's construction of his mandate, in my view, does not turn on the libertarian policies he favoured or his alarm about the growth of the federal government. The fundamental mistake lies rather in the

[45] Caplan, n. 15 above, p. 151.

[46] Ibid., p. 186.

[47] Fried, n. 8 above, especially pp. 17, 37, and 57.

[48] Ibid., p. 15. Joshua Schwartz, n. 10 above, properly counters Fried's grand declarations of ambition by observing that the SG is, to a large degree, a reactor rather than initiator in "legal policy". The SG depends, for example, on specific cases brought by other parties, and on adverse decisions in lower courts. Thus, it is misleading "to suggest that the Solicitor General is in a position systematically and comprehensively to plan and implement a unified legal philosophy" (p. 1094).

[49] Fried, n. 8 above, p. 17.

[50] Ibid., p. 37.

assertion that the basis for determining "what the law is" is his own philosophy—whatever that philosophy might be. Previous SGs, one could say, regarded themselves as bound by the legal record and the development of decisional law—whatever their own views—and thus saw the law as a constraint on their decision-making. Fried, however, was prepared to measure the law by the standard of his own philosophy. So he appears to say. And the appearance is supported by its conformity to a general tendency in his thought to rely on a conception of governance that turns institutional into personal relationships. This tendency shows itself in two ways: in his description of the SG's role and in his theory of law.

To begin, Fried construes the SG's relation to the President as though he was a White House staffer rather than head of an executive branch agency.[51] He refers to his job as that of "an advocate for the President".[52] And when he describes his duties, he frames them solely in terms of his personal loyalty to the president. This loyalty "is a moral attitude . . . [and] must be recognised as having intrinsic value, as being worth following for its own sake and in principle". Elaborating this idea, Fried says: "First we judge a thing or a person to be worthy of our loyalty, and then—and for that reason—subordinate our will to it . . . The President is entitled to loyalty conceived in just this way from those he appointed to high office".[53] This conception of loyalty could be construed as a positivist's construction of the chain of command, but by making it personal rather than institutional it goes beyond anything the positivist view actually requires. More importantly, Fried's construction leaves no room for the distinctive normative history of the SG's office, and makes no mention of the statutory origins of the office. The fact is that a large part of the SG's authority is derived from Congress, to whom loyalty is also owed. Thus, the analysis of loyalty is more complicated than Fried supposes.

I want to dwell on this last point because observers—and officials—often forget that executive branch administrators, while appointed by the President, operate under a statutory mandate and are accountable to the Congress for their conduct. This connection is even more compelling for administrators whose appointment requires Senate confirmation. Sometimes, indeed, Senate approval is conditional on specific declarations of intent or promises made by a nominee at the confirmation hearing. Consider the case of Eliot Richardson, when he was nominated by Richard Nixon to become Attorney General. His promise to the Senate not to fire special prosecutor Archibald Cox, except for just cause, was crucial to everyone's assessment of his obligations when Nixon ordered the firing. Even in the absence of such explicit declarations, the confirmation hearing should serve to remind appointees that they will occupy an office created by the Congress, with a more or less well formulated mission that stands independent

[51] Richard Neustadt draws this distinction in *Presidential Power and the Modern Presidents: the Politics of Leadership from Roosevelt to Reagan* (New York: The Free Press, 1990), p. 220.

[52] Fried, n. 8 above, p. 20.

[53] Ibid., p. 189.

of the wishes of the current occupant of the White House. The executive branch, in short, is not the President's fiefdom, with all political appointees owing personal fealty—and especially not the SG's office, given its normative history. In response, I suppose one could argue that, with Fried's explicit elaboration during the hearing of a positivist understanding of his role, the Senate implicitly agreed to his view in confirming him. However, after setting forth his supposition that the President had nominated him because of his philosophy, which would enter his judgement about what the law is on particular matters, Fried continued: "But certainly partisan political considerations have never entered into our judgments [in the SG's office], never should enter into our judgments, and I would never allow them to enter into our judgments".[54] So, there were legitimate grounds for confusion.

Fried resists the idea of conflicting commitments by adopting what he calls "the concept of the unitary executive", the locus of which is the person of the President, to whom all loyalty is owed.[55] In elaboration, he distinguishes two conceptions of loyalty, modelled on a classic distinction in the theory of democratic representation. The first, the mandate conception, would have the SG put himself so far as possible in the President's shoes when attempting to decide which position to take in a case, "trying to guess what the chief would do if he had the matter before him".[56] But the President does not have the SG's mastery of detailed and technical legal matters. What he has, rather, is "a general disposition about the law, about the Constitution, about how courts ought to work".[57] Therefore, the SG has to exercise his own judgement about the law, rather than guess what the President would have done.

This leads to the second, preferred conception of loyalty, which Fried calls Burkean—the trustee conception. Here the SG decides which position to take in specific cases by employing his interpretive skill "to make the best and most coherent whole out of his administration's projects and tendencies".[58] The President's "directives, pronouncements, hints, and actions" are the fragments out of which the SG constructs "a coherent morally and politically good whole".[59] So it is not a matter of guessing the President's mind but of constructing a position that the President could later embrace as his own. (Whether he would is another question.) Although the move is subtle, Fried hereby shifts from the President to himself. Moulding the fragmentary materials into a unified philosophy requires the interpreter to rely not only on "his own judgement" but "his own . . . values", in order to fashion the fragments into a coherent whole.[60] Thus, the skill that Fried needs, as an SG who does legal philosophy, is a capacity to bring his own values to bear in legal interpretation, albeit some-

[54] Caplan, n. 15 above, p. 151.
[55] Fried, n. 8 above, pp. 142–54.
[56] Ibid., pp. 173 and 188.
[57] Ibid., p. 191.
[58] Ibid., p. 173.
[59] Ibid., p. 191.
[60] Ibid., p. 173.

how on behalf of the president. In extension of this understanding of his assignment, if not entirely consistently, Fried had also said, at the confirmation hearing, that he had been chosen as SG because, from the beginning, Reagan had a sense of his (i.e., Fried's) philosophy—which was thus developed and known before he became the president's advocate and interpreter.[61] The President's expectation, accordingly, was that this philosophy would determine Fried's judgements of what the law is. By implication, Fried's claim is that the President delegated to him his own authority to interpret the law. That claim presupposes that the President has such authority to delegate, which is so only in the positivist description.

Let me emphasise that Fried presents this picture only in relation to the question of determining what the law is, where what he has in mind are the cases arising for decision before the Court. Presumably, though, the same picture applies to his construction of the SG's mandate, which is also a matter of (statutory) law. So we should understand Fried to have been gathering, as well, the many "directives, pronouncements, hints, and actions" that provided the materials for the administration's positivist view of the SG's office. To the extent that "his own values" figured in to this construction, the value of loyalty to the person of the President was central.

We should concede that the trustee conception—and his own skill in argumentation—could have given Fried a rhetorical advantage in debates with his colleagues in the Department of Justice. When disagreements occurred, Fried could argue that he had developed "a more faithful reading [than they] . . . of [the] administration's philosophy".[62] The most obvious difficulty is that Fried's rendering of the fragments necessarily downplayed the administration's moralistic streak. But let's leave that aside. The more significant difficulty is that the interpretive project, as Fried conceives it, is quite wrongheaded. At first, Fried's determination of what the law is employs fragmentary materials—hints, actions, directives, and pronouncements—derived solely from the President or the president's team. Then, when he needs a basis for making a coherent whole out of these fragments, he adds his own values—or substitutes his own philosophy for theirs. Astonishingly, in all of this, no reference is made to precedents, statutes, and other legal materials. Every determination of the law appears to be political, none legal. If Archibald Cox was known at the Department of Justice (according to Victor Navasky) as the SG "who couldn't see beyond the law", Fried would seem to be the SG who couldn't see beyond his philosophy. This posture not only neglects the institutional mandate of the SG and the normative history of the office; it rests on a faulty theory of legal interpretation.

In describing his interpretive task, Fried borrows from Ronald Dworkin's account of judicial decision-making. I think Fried is right to do so, but for a reason he does not mention and that does not fit his positivism. The reason is that the SG, as a guardian of decisional law, is more like a judge than any other kind

[61] Caplan, n. 15 above, p. 151.
[62] Fried, n. 8 above, p. 183.

of public official. Accordingly, the interpretive tasks of the SG should resemble those of a judge. So Dworkin's theory is apt. However, Fried gives it a personal cast that is insupportable—and that Dworkin, although initially drawn to such a view, no longer accepts. Dworkin was inclined for many years to regard the judge as a lone theorist, a Hercules, elaborating his vision of the law as a kind of personal achievement. More recently, however, he has stressed that each judge is a partner in a joint enterprise with other officials—past, present, and future—who together construct a coherent and common political morality.[63] From the beginning, Dworkin's concern has been about the exercise of discretion in a democratic polity. Do judges have a lawful, reasoned basis for decisions in controversial cases, or are their decisions at bottom arbitrary, having no firmer ground than personal predilections? If a judge's ruling is just one opinion among others, why should it take priority over whatever opinion emerges victorious in the legislative process? The positivist description of the judicial role heightens this concern. If law consists in discrete and readily identified official declarations, judges lack authoritative grounds for decisions in novel cases. In other words, in hard cases the law runs out. At such moments, a judge's exercise of discretion is, necessarily, uncontrolled by law. Citizens' legal rights and duties are the product of arbitrary invention, not reasoned discovery.

Dworkin rejects that possibility. Rather, every case where explicit law runs out poses a question of interpretation, requiring a kind of reasoned elaboration. Although the law is not literally pre-existing, it is constructed within definite constraints and may be said to be discovered. First, a judge's decision must fit the polity's previous political, including legal, commitments as embodied in statutes, case decisions, and other official pronouncements. Then, to the extent any indeterminacy remains, the judge extrapolates from these materials in the most morally appealing way. So in a hard case a judge faces a question of political morality, but the relevant morality is not personal. It is the political morality of existing law, that is, the emergent principles in the deliberations of officials over time and embedded in accepted legal practice. The question a judge asks is: What are the polity's deepest commitments, and what do they entail for concrete situations? To the extent that these background principles are themselves not sufficiently determinate, a judge's responsibility is to elaborate them consistently with what is clearly settled. This exercise aims at a coherent rendering of the polity's, not the judge's, basic commitments. The regulative ideal is integrity—the polity speaking in a single voice—which demands of any single judge such adjustments in his or her personal views as will accommodate the differing views of other officials. In that sense, each judge is asked to render impersonal moral judgements.

[63] The figure of Hercules appears in "Hard Cases", *Taking Rights Seriously*, n. 23 above, pp. 105–30, and in *Law's Empire* (Cambridge, MA: Harvard Univ. Press, 1986), p. 239ff. The communal conception of judicial decision-making emerges with Dworkin's introduction of the ideal of "integrity", which "takes the community itself as a moral agent" (p. 187). Integrity can conflict with—and properly takes precedence over—an individual judge's view of what would be just or fair (p. 176ff.).

The idea of impersonal judgement is sometimes rejected because of confusion about impersonal decision-making. It is thought that, if a decision does not depend on an official's personal attributes or values, the official is being required to act mechanically, without reflection or deliberation. Consequently, anyone in the position would be fungible with anyone else and would make the same decision. But it is a mistake to think that the only alternative to judgement based on personal conviction is like judgement by all. That would be the idea of an official as mere functionary, as though experience and expertise and reflection made no difference. Officials do not divest themselves of moral agency; both personal and impersonal judgements are made by persons. What distinguishes them are their grounds, the kinds of considerations taken into account, the point of view from which they are made—all of which leaves room for differences in judgement. Benjamin Cardozo captured the official's responsibility neatly when he said about judicial decision-making: "The thing that counts is not what I believe is right. It is what I may reasonably believe the person of normal intellect and conscience would reasonably look upon as right".[64]

Applying this thought to the SG, we can see how the determination of "what the law is" fuses with moral and political argument. Suppose Fried had been a modern-day Plato and the construction of his mandate was logically entailed—or at least well supported—by a comprehensive political theory developed by him. Suppose, further, that the theory was brilliant and morally compelling. As such, it could form the basis of a political platform. But it would have no legal standing whatever—unless the theory could claim grounding in settled law. In other words, it would have no standing unless the very last thing one would say about it is that it was Fried's theory. The relevant "values" are not those Fried brings to bear when he does legal philosophy—his own. They are the general moral principles that existing law presupposes by way of implicit justification. Explicit law, we could say, is only the more evident aspect of the body of norms rooted in the moral conventions and understandings of community members. Accordingly, when the SG—or the SG's boss—makes a judgement of political morality, and therefore of how decisional law should develop, the judgement will be only as good as the arguments that could be mustered showing that it would be part of the soundest construction of decisional law, in light of the community's background morality. Clearly, a philosophy reflecting alienation from forty years of legal development would not be able to make that claim.

The naturalist alternative

We have, then, reasons for rejecting the positivist orientation to law, and especially Fried's application of it. The naturalist, I would say, is more faithful to the social work actually done in legal practices and more appreciative of their real

[64] Benjamin Cardozo, *The Nature of the Judicial Process* (New Haven: Yale Univ. Press, 1921), p. 89.

achievements or failures. In particular, legal institutions are congeries of purposive relationships, not just hierarchical structures of rules. Fuller articulated this point by distinguishing a structural view of institutions from an interactive view, offering an analogy with language. To convey meaning, language must have definite forms and rules, which thus regulate our communicative conduct. Respect for these constraints is necessary if we wish to be understood at all. But it is a mistake to focus exclusively on the formal structures and forget the purposive activity that sustains them—and that may push them in new directions. Sometimes effective communication requires us to violate established rules, creating new forms of expression. So with legal interaction. To accomplish their goals, participants in legal agencies must follow the forms set by previous interactions, but new interactions could move beyond them. As a result, formal rules specifying the powers and duties of a role must be treated in a qualified way, not insisted upon regardless of what happens to the aims of the agency.

Here, too, is a lesson about rules. As the *Tedla* Court instructed us, rules (or rule-formulations) are imperfect or incomplete, commonly by being under- or over-inclusive in relation to their rationales. Consequently, intelligent rule following requires second-order competencies, such as the ability to discern the purpose of a rule and to make appropriate adjustments. When we are talking about a complex of rules, designed for sophisticated ends, the need for second-order competence multiplies. And as sustained adjustments occur, new rules and expectations develop. So, we do not fully understand an agency or institution without getting at the basic attitudes and dispositions it requires, and without knowing whether a given structure of rules fosters or inhibits them. Only in this way do we connect to participants as moral agents with moral aims, such as the SG's fiduciary duty to the law. As I have indicated, this duty belongs at the centre of the naturalist description because the SG's practice, so conceived, promotes the rule of law.

The fiduciary duty is exhibited in the patterns of expectation that govern the SG in interactions with the Supreme Court. The SG submits a brief to the Court, either as one of the parties or as amicus curiae (friend of the court), in almost every case where the government is implicated. In these submissions, the Court expects the SG to be a counsellor to the Court, the legal conscience of the government—not, simply, an agent of the current administration. That means the SG looks beyond the platform of the administration as well as beyond the interests of the immediate parties in a case, to guide the Court in taking the long view, toward what we have called the orderly development of decisional law. In this capacity, the SG's job is as much to protect the Court as to persuade it. Correspondingly, in discussions with administration officials, the SG should give voice to the views of the Court and deliberations within the legal community. That is what it means to be a guardian of the law.[65]

[65] Historically, an amicus brief was regarded as "friendly" because it was construed as an aid to the Court, for example, by highlighting a legal point that might be slighted or might escape the Court's attention. In this way, the Court was protected against mistake or wrongdoing.

With the Court relying so heavily on the SG, the SG has correspondingly weighty responsibilities to the Court. The SG can be effective only if the justices have confidence in the SG's professional integrity and legal judgement. Whether they have such confidence is in great part a function of the evident commitment of the SG to professional norms and the skill with which the SG assists the justices in meeting what they understand as *their* responsibilities. Thus, the relation involves an appropriate reciprocity, which can go awry if the SG tries to get the Court involved in ideological squabbles. Skill in the lawyerly craft is essential to maintaining credibility and shows itself in a number of concrete ways. One is to take seriously the legal issues in a case and keep them distinct from partisan political stances. Ideology predominates when the SG's brief to the Court downplays lawyer's issues and becomes a position paper on public policy. More generally, the confidence of the Court in the SG is based on the SG exercising self-restraint, for example, by refraining from asking the Court to address a major issue without having engaged in careful intellectual preparation. An argument for reversal, to take one instance, becomes compelling only when it is preceded by thoughtful lower-court decisions and changes in the intellectual climate of the legal profession, showing serious reconsideration of the Court's precedents—not when the President's political agenda calls for it.

In some passages in his apology, Fried acknowledges these points. He recognises that a different conception of the SG's office had prevailed before the Reagan years. The office was "disconnected" then from the administration in power, was staffed by a "special breed" of career lawyers who produced "high-quality legal work, scrupulously fair to facts and law", and who upheld "an ideal that entails a kind of regularity, objectivity, and professional technique apart from—maybe even above—politics".[66] Later, Fried concedes that "even in constitutional cases, precedent and analogy are the stuff of legal argument, and . . . legal argument is what moves the Court—or moves it when all involved are doing their work right".[67] Even at his confirmation hearing, Fried expressed this traditional conception in part of his testimony. He said: "Nothing would be more important to me than to maintain that sense of confidence which I believe the Supreme Court has always had in the Office of the Solicitor General. That the Supreme Court can believe that the work that comes from the office represents the most objective, the most accurate, and the fairest presentation of the issues before it".[68] And, in his book, he recounts instances of inappropriate interference to which previous SGs had been subjected by their bosses or other executive branch officers. Chief among these is the effort of Joseph Califano, when he was HEW Secretary, to orchestrate pressure on SG Wade McCree to change the government's brief in the *Bakke* case to a position more favourable to affirmative action. (Fried is temperate in his condemnation of Califano's blatant disrespect for the professional judgement of the career lawyers. He refers

[66] Fried, n. 8 above, pp. 36–8.
[67] Ibid., p. 66.
[68] Caplan, n. 15 above, p. 151.

to the shouting match between Califano and the lawyers as a "scene lack[ing] decorum".)[69] Fried also recognises that the ideal of the rule of law, to which he says he is strongly committed, "holds that [people] of intelligence and reasonable good will can come to a fair measure of agreement about what the law is".[70] But surely he had to know that such agreement cannot be expected if a determination of the law is based on the SG's personal philosophy, especially (as I have emphasised) a philosophy based on a rejection of forty years of legal history.[71]

The contradiction in Fried's position is evident in the hope he expressed when he first took office: "What I wanted was to establish my authority as head of the office, but in a way that respected the tradition that career lawyers should be able to work for a conscientious political chief even without sharing his political commitments".[72] It was precisely Fried's (and the President's) "political commitments", however, that undermined his authority as SG, by sabotaging the institutional disposition to respect and defer to the career lawyers. On this point, were it not tangential to our present concerns, we could fruitfully examine Fried's political commitments more closely (even drawing on his other writings) and spell out the implications for his reading of the law. Here let me say just a word about this connection as it appears in his apology. Most importantly, in a manner that is characteristically libertarian, Fried regards law and liberty as antitheses. "Laws restrict liberty", he says, even though liberty cannot exist without laws.[73] The reason is that, to exercise their liberty, citizens need protection from the depredations of others, and law provides that protection. In this regard, Fried is in accord with Austin, as well as Jeremy Bentham, who regarded the existence of law, while indispensable to human society, as at best "a necessary evil".[74] But the implication, at the limit, is that what is most valued by individuals is a life completely without legal constraints. (Anthony Flew characterises libertarianism, which he defends, as a doctrine "opposed to any social and legal constraints on individual freedom".)[75] I am confident that Fried

[69] Fried, n. 8 above, pp. 199–201. For Califano's own account, see *Governing America: An Insider's Report from the White House and the Cabinet* (New York: Simon and Schuster, 1981), pp. 236–43.

[70] Fried, n. 8 above, p. 59.

[71] An illuminating contrast would be the approach of the SG's office when it urged the Court, in *Brown v Board of Education*, to overturn state segregation laws. Attorney General Herbert Brownell understood the importance of having bipartisan support for the government's brief in the case, that is, from both the Truman and the Eisenhower administrations. Further, the crucial precedent to be overturned, *Plessy v Ferguson*, had been handed down more than fifty years before. And, most importantly, the government's argument in *Brown*, despite its being in some respects a radical departure from the past, was constructed on the underlying logic of a series of recent "separate but equal" decisions, which effectively placed black students in white institutions, since the black institutions were not equal.

[72] Fried, n. 8 above, pp. 38–9.

[73] Ibid., p. 60.

[74] Bentham's language is unwavering: "[Law] may be a necessary evil, but still at any rate it is an evil. To make law is to do evil that good may come": *Of Laws in General* (H.L.A. Hart (ed.), London: Athlone Press, 1970), p. 54.

[75] *A Dictionary of Philosophy* (London: Pan Books, 1979), p. 188, quoted by Will Kymlicka, *Contemporary Political Philosophy* (Oxford: Oxford Univ. Press, 1990), p. 146.

does not mean to go that far, but his dominant worry about "government aggrandizement" is not sufficiently fine-grained to capture how acts of government enhance freedom rather than diminish it.

Fuller would have agreed that liberty takes priority among political values, but he would have added that law is not inherently inimical to it. Fuller belonged to a tradition of liberal theorists, beginning with some suggestive remarks by Locke, who regarded law and liberty as having a special affinity, because law—properly fashioned—enables liberty to be realised. This happens not (merely) by restraining others from interfering in individual choice but by providing appropriate institutional mechanisms that enable individual choice to be socially effective.[76] The importance of this positive conception of liberty, in the present context, is that Fuller's view is arguably closer to the one adopted by the Supreme Court beginning in *West Coast Hotel v Parrish* and continuing through subsequent cases. Fried recognises the importance of the 1937 decision, in which the Court struck down long-standing barriers to government regulation of the economy. He remarks that it was more than a political decision; it was an intellectual decision, reordering the whole body of constitutional jurisprudence to accord with a new understanding of fundamental constitutional values. Further, it was necessitated by "developments that made the economy indisputably national" and that "displayed the economy as inextricably . . . made up of a web of relations which government in part creates, in part sustains, always meddles in".[77] While not exactly, shall we say, applauding these developments, Fried insists that cynical doubts about the necessity of the jurisprudential transformation—expressed, for example, by Robert Bork—are unwarranted. Fried also recognises that any principled rejection of forty years of settled law threatens to require repudiation of the decision in *Brown*. But such wholesale repudiation, he says, is bound to be "anticonstitutional and thus lawless".[78] And he criticises Bork again on this score—lending credence to the defeat of Bork's nomination to the Supreme Court.[79] In Fried's view, to the contrary, *Brown* has to be taken as a "Himalayan fact" that took "the largest step since Reconstruction in dismantling legalized racism and apartheid". Yet, curiously, Fried adds that *Brown* helped to make "our system of individual liberty a model for the whole world".[80] Appealing to *Brown* as a criterion of any acceptable theory of constitutional adjudication is one thing, and not surprising even for a theorist inclined to align himself with challenges to the post-New Deal presumption in favour of a regulatory state. But it is surprising to see the accent put on liberty rather than equality. If liberty is the preferred value, it is certainly not the libertarian conception that covers this case. What Fried fails to mark is that the institutional reordering required by decisions such as *West Coast Hotel* and

[76] See Winston, "Legislators and Liberty" (1994) 13/4 *Law and Philosophy* 389–418.
[77] Fried, n. 8 above, p. 69.
[78] Ibid., p. 18.
[79] Ibid., p. 65.
[80] Ibid., p. 70.

Brown involved a reconceptualisation of liberty. It came to be understood less in terms of the absence of external impediments and more in terms of the meaningful exercise of basic human capacities or powers in controlling one's own fate—where "one's own" is collective as well as individual. Fried's libertarianism, I believe, is not consistent with that historical transformation.

Conclusions

If the Reagan team had succeeded in redefining the mission of the SG's office, they could not then have expected the SG to continue exemplifying a fiduciary duty to the law. Nor could they have expected the SG to retain the confidence of the Supreme Court and to continue exercising the influence on the law's development that the office had wielded traditionally. The Reagan administration, it seems, wanted things both ways. They wanted to press their political agenda, but they also wanted to retain the authority of the office, which depended on the informal norms I have identified. So perhaps it is closer to the truth to say that the Reagan administration wanted to take advantage of the prestige of the office, to serve an agenda that they had to know would undermine that prestige. In the end, their attempts are best understood, not as an effort to establish an alternative conception of the office, but an effort at political manipulation of the traditional conception. Their strategy was necessarily short-term and opportunistic.

Since Rex Lee's and Charles Fried's superiors were not concerned about the integrity of the office, it fell to the SGs themselves to resist the political pressures. The basis of resistance was an understanding of what the agency aspired to be and do. It was thus a standpoint within the agency, rather than external to it, that provided the resources for critical reflection. (Of course, since an agency never perfectly achieves its ends, it has always a double character—its real and ideal aspects.) What the SGs in this case took that understanding to be, I have assumed, was open to dispute. In principle, the conception of any public office can be contested. But the principal lesson of this inquiry is the need for caution in what we think the contest looks like. It may be thought, for example, that each occupant of an institutional role is free to select any of the variety of competing conceptions of that role available in the political culture. In an abstract sense, that may be so. But a conception may be "available"—and a role occupant "free" to adopt it—only with great cost to the mission of the office in which the role figures. Not everything is in play. The constraints that preclude the choice of a conception outside the currently available ones—past practice, prevailing expectations, the conditions of institutional integrity—may also authorise only a single choice among "competing" conceptions.

15

Administrative Policy-making: Rule of Law or Bureaucracy?

HENRY S. RICHARDSON*

I. INTRODUCTION

The vast increase in the administrative apparatus of modern governments over the last few centuries has posed many questions for which the political theories of republican and democratic government that we have inherited from the eighteenth century leave us without ready answer. Not least of these concerns the rule of law. How should our understanding of the rule of law take into account the proliferation of administrative agencies? What new challenges does the existence of these agencies pose to the rule of law? And what do these problems reveal about the relation between democracy and the rule of law?

I am aware that on certain conceptions of legality, the rule of law and democracy are inherently connected, perhaps because legality is thought of as intimately tied to legitimacy, and legitimacy, in turn, depends on democracy.[1] As I will be using the term "rule of law", however, the question of the relation between the rule of law and democracy is more open than this. While I agree that the legitimacy of laws does depend on democracy, I think that there is a thinner understanding of the rule of law which does not carry with it all of the commitments of legitimate legality. This narrower, traditional interpretation of the rule of law may be summed up under three headings: generality, predictability, and regular process.[2] The laws must be general, so as not to single out any individual person or firm for special punishment or favours. Bills of

* I am most grateful to David Dyzenhaus both for the opportunity to reflect on these issues and for detailed comments. A version of this essay was presented at the annual meeting of the American Political Science Association in September 1998. I learned a lot from Stephen M. Young's commentary on that occasion. A later version was presented at the Kennedy Institute of Ethics at Georgetown University, where Tom L. Beauchamp, George Brenkert, and LeRoy Walters offered particularly useful criticism. I also thank Thomas Christiano and Jody Freeman for instructive written comments.

[1] See, e.g., David Dyzenhaus, *Legality and Legitimacy: Carl Schmitt, Hans Kelsen and Hermann Heller in Weimar* (Oxford: Oxford University Press, 1997).

[2] My account, here, is an amalgam of William Scheuerman, "The Rule of Law and the Welfare State: Toward a New Synthesis" (1994) 22 *Politics & Society* 195–213, and John Rawls, *A Theory of Justice* (Cambridge, Mass.: Harvard University Press, 1971), section 38.

attainder, which impose punishments on named individuals, are not to be considered law. General law must be uniform across persons, treating like cases alike. While generality will help with predictability, it is not sufficient for it. Predictability is required in order that citizens be able reasonably to rely upon the law in planning their conduct. To this end, law must be promulgated in a public way by regular procedures. In order that laws be accessible to citizens' planning, they must be promulgated in advance of their sanctions coming into effect: there may be no *ex post facto* laws, at least not ones imposing punishments.[3] It is unfair to impose punishments on people who, when they acted, had no basis for understanding their behaviour as criminal. Imposing punishments in the absence of any law declaring behaviour illegal is similarly unfair. A final prerequisite of predictability and fairness is that laws must not be unduly vague—no vaguer, at least, than the way I have formulated this requirement! Generality must not become meaningless abstraction. Finally, the requirements of regular process ensure, at the second order, that existing laws are applied fairly and with at least a minimum of consistency. While ways of institutionalising regular process differ, in general they support open procedures bound by rules of evidence, which allow for affected parties to present their arguments and challenge their opponents. While satisfying the requirement of regular process seems not to require democracy, it does require a somewhat independent judiciary. It depends upon judicial officials who are subject to rules that help ensure their impartiality and shield them from bribery and other undue attempts at influence. These three requirements, then, of generality, predictability, and regular process make up the narrow, traditional notion of the rule of law that I shall use here. While the precise institutional details of their interpretation will of course vary, these three requirements hang together. All are directed towards regularity in the content and application of the law.

Understanding the rule of law in this way, I shall examine the challenges posed by the growth of the administrative state. I shall argue that while there are serious questions about the legitimacy of the administrative state, inadequacy of the rule of law therein is not generally one of them. Whereas the ideal of the rule of law speaks to the content and application of laws, the real problem with the administrative state concerns the *generation* of laws—specifically, it concerns how they may be democratically generated. The rule of law, however, as I shall argue, is silent on the key questions pertaining to the democratic control of administrative action. We shall see, then, that the rule of law, while a necessary support of legitimate democracy, is hardly a sufficient guarantor thereof. Yet examining the way that the rule of law might work out at the administrative level is a very useful means for us, as we theorise about the democratic process, to think about what democracy requires under present-day conditions. In par-

[3] In American constitutional law, the ban on *ex post facto* legislation has, for centuries, been limited to the criminal law. See, however, Justice Thomas's concurring opinion in *Eastern Enterprises v Apfel*, handed down 25 June 1998 (No. 97–42).

ticular, as we shall see, different proposals for how to reconcile administrative power with democracy invoke different understandings of the rule of law.

In what follows, then, I shall argue, first, against Theodore Lowi and William Scheuerman that the modern administrative state is fully compatible with the rule of law.[4] Lest you think me complacent, I shall show, secondly, that contemporary administrative power does pose serious normative problems, wholly independent of the content of the programmes being administered. These are problems about the compatibility of such vast administrative power with democracy. I shall argue, thirdly, that making administrative power compatible with democracy does require that the rule of law extend to the agency level. So the real question, which I shall take up last, is what the rule of law must look like at that level. Thus, I shall be arguing that the real problem is not *whether* the rule of law applies at the agency level—for, as I shall argue, it already does—but how we must *interpret and adapt* the ideal of the rule of law at the administrative level if we are to reconcile the modern state with democracy.

II. MODERN ADMINISTRATIVE STATES SATISFY THE RULE OF LAW

Is the rule of law really compatible with vast administrative power? There is a serious question about this, as Scheuerman has effectively pointed out in an informative article on the subject.[5] If administrative agencies simply put into effect the laws passed by the legislature, without altering or amplifying them in any way, they would not have much power. They would be mere instruments of the legislature, doing its bidding. What constitutes the power of contemporary government agencies is that they are not limited to such a mechanical role. Rather, agencies acquire important substantive power in two main ways: because legislatures explicitly delegate it to them or because the legislation they must implement is vague and inconclusive, leaving the agency little choice but to settle broad policy questions. The vast administrative power that has arisen both from explicit delegation and from the implicit delegation by vagueness threatens the rule of law. While most of my arguments in what follows would, I think, apply equally to agency power arising from either of these sources, I shall concentrate on the problem of statutory vagueness.

Statutory vagueness is endemic in modern governments, and has long engendered a sense that their legitimacy is in crisis. It will be our task to try to describe the problem with legitimacy more precisely than it has been hitherto. To begin this task, and for the sake of completeness, let me put some examples of this sort of statutory vagueness before us. As you will see, they exist across the whole range of government action:

[4] My targets, here, are Theodore Lowi, *The End of Liberalism: Ideology, Policy, and the Crisis of Public Authority* (New York: Norton, 1969), and Scheuerman, "The Rule of Law and the Welfare State", n. 2 above.

[5] Scheuerman, ibid.

(1) *Vagueness in the context of explicit delegations to paradigmatically regulatory agencies*: the Federal Trade Commission Act of 1988 required the FTC to "prevent persons . . . from using unfair methods of competition". The Federal Communication Act of 1988 mandated that the Federal Communications Commission regulate the airwaves "as the public convenience, interest, or necessity requires".[6]
(2) *Vague resolutions of environmental issues*: the Clean Air Act Amendments of 1977 declared that the air quality in the USA should not degrade. It was left to the Environmental Protection Agency to work out what this might mean.
(3) *Thoughtless extensions of entitlement programmes*: in 1972, a congressional staffer drafting the legislation enabling the Supplemental Security Income programme for the disabled inserted a twenty-six-word parenthesis that extended the programme to children. Since the Bill's unemployability standard of disability could not be applied to children, it was left to the Department of Health, Education, and Welfare (as it was then) to work out how to extend the programme to children; the Department did so, developing what became a three-billion-dollar-a-year programme of disability payments for children.
(4) *Imprecise retrenchments of entitlement programmes*: by 1996, the conservatives in Congress had come to the conclusion that disability payments to children had gotten out of hand, breeding corruption. The retrenching law was, in turn, vague, leading to many more children's benefits being terminated than had been anticipated.[7]

As these examples indicate, the amount of agency discretion afforded by statutes can be staggering. As they also begin to illustrate, statutory vagueness can be found in widely varied substantive areas of the law. And the reasons for statutory vagueness are just as varied. Sometimes, as in the case of extending disability benefits to children, innocuous wording simply slips by the legislators as part of a complex package. On other occasions, as with the FTC and FCC acts, vagueness is a means of deliberate delegation of power. At other times, vagueness can be an indispensable grease for the wheels of compromise, albeit one that will come around to cause other problems later on. Finally, as in the case of the Clean Air Act Amendments, vagueness can reflect legislative awareness of the technical difficulties of attempting to be any more precise.

It should be obvious that, taken by themselves, provisions as vague as these fail to live up to the requirements of the rule of law. Industries concerned with their pollutant output, families with disabled children, and business competitors hardly have been given a firm basis for planning their activities by these statutes. This gives rise to the spectre of unfairness and arbitrariness in their application.

[6] Taken from Mark Seidenfeld, "A Civic Republican Justification for the Bureaucratic State" (1992) 105 *Harvard Law Review* 1551–76, n. 1, citing 16 U.S.C. §824e(a) and 47 U.S.C. §303.

[7] Robert Pear, "After a Review, 95,000 Children Will Lose Cash Disability Benefits", *New York Times*, 15 August 1997, p. A1.

On Scheuerman's analysis, these threats to the rule of law have too long been seen by political theorists as intrinsically bound up with an interventionist state. Hayek, as he reminds us, saw the tendency towards vague legislation in modern states and concluded that the rule of law was incompatible with the sorts of legislative intervention that these states attempted. The only way to maintain the rule of law, Hayek insisted, was to roll back the activities of government. And the rule of law must be maintained. "Law in its ideal form", Hayek writes, "might be described as a 'once-and-for-all' command that is directed to unknown people and that is abstracted from all particular circumstances of time and place and refers only to such conditions as may occur anywhere and at any time".[8] In this way, law in its ideal form can avoid the kind of limitation on liberty that comes about when one person's command effectively removes another's power of choice. This requirement of abstraction Hayek further interpreted as implying that "law" is limited to "general rules that apply equally to everybody".[9] This, in turn, places a special burden, in Hayek's view, on any purported element of law that makes any distinctions among citizens. Yet this, on Hayek's analysis, is what an interventionist state aims to do: it aims to take from some citizens and give to others. Hence, paring back to a libertarian state is not just an important way of reducing the decision-making burden on legislatures; it is also conceptually necessary to having laws that are uniform in their application to all citizens.

While I agree with Scheuerman that statutory vagueness poses a crucial problem of legitimacy, I do not think that it is best to understand this problem as one of incompatibility with the rule of law. In laying out why not, it will be useful to start with what is wrong with Hayek's interpretation of the rule of law. There are two problems. The first is that he misidentified the source of the problem of vagueness. As my examples begin to indicate, it is not only redistributive programmes which have this problem. Further, as Scheuerman well argues, there is a strong *prima facie* reason to think that statutory vagueness will redound to the benefit of the economically powerful, for they have the money to influence agencies in whatever leeway the statutes leave.[10] Accordingly, it is odd to cast statutory vagueness as a creature of the friends of redistributive justice. Rather, as I have briefly indicated, the sources of vagueness are more varied and more boring than this. The second problem with Hayek's analysis is that his interpretation of the generality requirement is absurd. All law makes distinctions among citizens. To take just some of the favourites of the libertarians: property law distinguishes between owners and trespassers, contract law between promisors and promisees. Hayek's understanding of the generality requirement, if

[8] F. A. Hayek, *The Constitution of Liberty* (Chicago: University of Chicago Press, 1971 [1960]), pp. 149–50.

[9] Ibid., p. 153.

[10] As David Dyzenhaus notes in his Introduction to the present volume, this point does not apply as forcefully in the case of explicit delegations, many of which (at least in the past) have created agencies intended to aid the disadvantaged.

enforced, would reduce law to vapidity. Or, to put the point in a more conciliatory mode: *any* provision of law, even a redistributive one, may be put in a fully general way that, in form, makes no distinction among citizens. That is, any provision may be put in the form, "for all citizens, if A then B". If Hayek is insisting that no description of the citizen may appear in the antecedent, A, this is, again, the absurd position. It would dictate that there could be no law saying, "for all citizens, if a citizen owns a piece of land, he may keep unwanted people off it". Yet if citizens may be described in the antecedent, then there is no basis in the idea of generality for distinguishing between "if a citizen owns a piece of land" and "if a citizen earns more than $200,000 a year". Now, I said that *any* provision of law may be put in this general form; but of course I mean any provision which does not truly violate generality by naming a particular individual. (The possibility of rigged general descriptions, which do not syntactically name a particular individual, but which are crafted so as to pick out a single individual, does pose a theoretical problem, but may more easily be dealt with in practice.) The absurdity of Hayek's interpretation, then, reminds us of what the generality requirement really comes to: a ban on laws aimed at particular individuals, as opposed to classes of individuals (owners, promisors, the rich) who are, by the legislators, treated as relevantly "alike".[11]

With this clarification, we may turn to the central reason why statutory vagueness is no threat, by itself, to the rule of law. The reason is that the administrative agencies of modern states have evolved ways of making policy that themselves satisfy the requirements of the rule of law. Thus, while agencies do often make decisions about individual cases—the IRS comes to mind—they do not operate solely on the basis of the statutes whose vagueness we have noted. Instead, they make policy by promulgating administrative rules. In the USA, the Administrative Procedures Act of 1946 made provision for what has come to be called "notice-and-comment rulemaking", a set of procedures in terms of which the most important intermediate policies are set out. In the case of the extension of disability payments to children, for example, the twenty-six-word parenthesis in the statute, which was entirely vague as to what disability might mean in children, became thousands of pages of detailed rules promulgated in the Code of Federal Regulations. A similar process is found in many other countries. In Germany, the constitution of 1948 makes provision for what are called *Rechtsverordnungen*, which serve just the same function. In both countries, rules have the force of law.

Not only that: because these administrative rules are *not* vague, they provide compliance with the rule of law that would be missing from the statutes that

[11] In *The Constitution of Liberty*, Hayek seems to me to have been caught between contradictory commitments. On the one hand, he rails against "the emasculated concept of the merely 'formal' *Rechtsstaat*" which lost sight of "the ideals which inspired the liberal movement" (p. 484 n. 35). On the other hand, in building a conception of law in support of his own libertarian ideals, he strives to build it into requirements—such as that of generality—that purport to be merely formal. As I argue in the text, this does not work.

they interpret, taken alone. I do not claim that all elements of vagueness and ambiguity are removed by administrative rules; still, the relationship between rule and statute in the disability case is quite typical. The rules spell out in great detail how a given, perhaps quite vague, statutory provision is to be interpreted by the agency. While some questions of interpretation will inevitably remain, it would be more apt to complain of density and complexity in the regulations than of vagueness. Perhaps the agencies sometimes go overboard in spelling out details, and produce regulations that make for less than optimally predictable law, because too opaque to the citizenry. Nonetheless, a certain amount of density and complexity seems to go with the territory of the law. Whereas the vague statutes cited in my examples do seem, taken by themselves, to fail in carrying out the rule of law, when these same statutes are taken together with the administrative rules that fill out and interpret them, they succeed. They present general rules, promulgated by a regular process, that serve as an adequate basis for predicting the legal consequences of conduct. Accordingly, they meet the three requirements of the rule of law: generality, regular process, and predictability.

III. ABDICATION OF LEGISLATIVE CONTROL

Against what I have been arguing, you may object that I have stretched the meaning of the term "law" beyond its original meaning. "Laws", you may argue, "are what is made by the law-makers, that is, by duly constituted legislatures. No other provisions can count as laws, however they may dress themselves up. The point of the rule of law", you may explain, "is to insure that government is controlled by the democratically elected legislators. When these officials delegate or abdicate their responsibilities by passing vague legislation, then the democracy loses control of the law, and there is no longer a true rule of law".

Now, I think that an important truth underlies this objection, but that the objection distorts it. Remember that I stipulated at the outset how I would be interpreting the idea of the rule of law, namely in terms of the requirements of generality, predictability, and regular process. I also gave capsule accounts of the rationales for each of these requirements. These had to do with avoiding basic unfairness and arbitrariness and allowing citizens a firm and regular basis for planning their activities. They had nothing to do, explicitly, with democracy. All these conditions could be perfectly well satisfied by a constitutional monarchy. It seems, then, that administrative rule-making not only comports with the defining requirements of the rule of law, it also satisfies the underlying goals of this ideal.

I do not mean to stretch this point too far. As David Dyzenhaus has noted, much danger may lurk "under the cloak of legality".[12] If we imagine moving

[12] David Dyzenhaus, *Legality and Legitimacy*, n. 1 above, p. 217.

away from a democratic setting altogether, perhaps to a fascist one, the possibility of maintaining the rule of law as I have interpreted it may show that rule to be a sham. There are several possibilities, here. One is that switching to a fascist context would force us to be more definite about vague clauses pertaining to "regularity", "impartiality", and "regular process" in the defining conditions of the rule of law. Another is that we must simply recognise that the rule of law is an incomplete and subordinate ideal, subservient to the broader and more important ideals of democracy and justice. For the sake of conceptual clarity, in any case, I continue to want to use the term "rule of law" to cover the fairly formal set of conditions I went over at the outset.

What is true about the objection, though, is that the tremendous transfers of power to administrative agencies, which call upon them to fill out almost all substantive details by writing administrative rules, do amount to a failure of *democracy*. Here it matters somewhat whether the transfer is purposeful or not. Many broad delegations intended to empower independent regulatory agencies reflect a democratic judgement that power should thus be transferred. In those cases, the vagueness of the statute corresponds to the breadth of the intended transfer of discretion. In many other cases, however, statutes are either unintentionally vague, through thoughtlessness or imprecision, or else vague because nothing more precise could garner agreement. For these latter sorts of case, certainly, it is disingenuous to suggest that the democratic intention of the public was simply to transfer the corresponding discretion to the agencies.[13] In both kinds of case, the resulting scope for administrative rule-making poses a threat to democratic legitimacy—in the case of purposeful delegations, a *prima facie* threat; in the case of unintentional imprecision and vagueness due to stalemate, the threat is severe.

If the resulting scope for administrative discretion is compatible with rule by the people, we lack an adequate way of understanding how this might be. In this respect, our understandings of administrative rule-making are in a different state than our understandings of the creation of legislation by representative legislatures or its review by the judiciary. To be sure, both representative government and judicial review move us some steps away from a pure, direct democracy. Recent books by Bernard Manin and Philip Pettit remind us of the original distinction between democratic ideas and the republican ideas of separation and checking of powers.[14] Nonetheless, the centuries have accustomed us

[13] With a public choice theorist's rather minimal expectations for popular sovereignty, Jerry L. Mashaw has, in fact, used the point that vague statutes are passed by legislatures accountable to the electorate to argue that statutory vagueness poses no threat to legitimacy. See especially his *Greed, Chaos, and Governance: Using Public Choice to Improve Public Law* (New Haven: Yale University Press, 1997), pp. 138–40. If having a chance to "vote the bastards out" would suffice for democracy, this would be an adequate response to the threat to legitimacy posed by vague statutes. This conception of democracy, however, is too thin.

[14] Bernard Manin, *The Principles of Representative Government* (Cambridge: Cambridge University Press, 1997); Philip Pettit, *Republicanism: A Theory of Freedom and Government* (Oxford: Oxford University Press, 1997).

to their combination, and most normative accounts of democracy accommodate the resulting attenuations or impurities from the outset. It is not so with administrative rule-making. For too long, we have been lulled into theoretical complacency by the thought that what the agencies did was simply technical, simply a way of finding efficient means to the ends set by legislators. In the case of broad and vague delegations, however, it is obvious that the agencies must also set ends.[15] Hayek, to be sure, was not complacent about this; but, as we have seen, he cast his complaints in an exaggerated form that, I conjecture, has deprived him of listeners. What seems plain, if we look at matters squarely, is that the bulk of the really hard political decisions, the tough decisions requiring the specification of vague ends and the compromising of competing ones, goes on in the administrative agencies and issues in administrative rules. And we have no good account of how it is that the people is thereby ruling.

I stress that it is an adequate normative account that I believe we lack. Without a more developed normative account of how the demos *should* be exercising control over administrative rule-making, it is hard to assess how things now stand. According to current notice-and-comment procedures, agencies publish proposed rules in the *Federal Register*, collect sometimes thousands of comments from the public, and painstakingly prepare answers to them. Under certain imaginable normative conceptions, it will turn out that these steps are perfectly adequate to protect democracy. I myself am doubtful that these current procedures are adequate to legitimate agency policy-making. It is hard to say whether these procedures are adequate bulwarks of democracy, though, until we have better articulated what democracy requires of administrative rule-making. Though the very question sounds boring, it is of vital importance to the legitimacy of our governments.

My constructive means of making progress on this question, here, will be to explore alternative detailed ways of working out the rule of law at the administrative level. Before I come to that, however, it will be important to remind ourselves that, although insufficient for ensuring democracy at the administrative level, the rule of law is certainly a necessary condition of democracy at that level.

IV. THE RULE OF LAW IS ESSENTIAL TO DEMOCRACY

It is easy to imagine how, absent the constraints of something like the Administrative Procedures Act, administrative rule-making could become merely a cloak for special interests and entrenched powers. If agencies could make policies without formulating them in publicly promulgated rules, then, unless democratically-elected legislatures could somehow take back the policy-making reins, democracy would be a sham. Writing almost ten years ago, Hernando de Soto contrasted what he called the "democratic" rule-making in the USA with the way administrative rules were then made in Peru:

[15] I argue this point more fully in "Democracy and Administrative Rationality" (in progress).

> "Peru is considered a democracy because it elects a president and a parliament. In the five years after an election, though, the executive branch has been known to make 134,000 rules and decrees with no accountability to the congress or the public. After elections, no ongoing relationship exists between those who make decisions and those who live under them".[16]

Admittedly, to make a fair and full comparison between Peru and the USA, we would have to discuss the degree to which the American President makes policy by executive order. My point, however, is not about these particular countries but about the requisites of democracy. On this matter, the purported contrast between the two countries suffices to make the argument obvious. Democracy requires that the people who live under laws have a substantial voice in their generation. Vague statutory delegations provide no input into many of the most serious policy questions within their rubric. Hence, if the people are to have a substantial voice in the generation of the rules under which they live, they must have a substantial voice in the making of administrative rules. Without an Administrative Procedures Act or something like it, though, such a voice is impossible. As the (purported) case of Peru shows, it is quite possible for special interests to take control of administrative rule-making in the windowless conference rooms of national bureaucracies. Therefore, extending the rule of law to agency rule-making, the way that the Administrative Procedures Act did, is a necessary condition of democracy.

This conclusion has backward-looking implications for legislation. These emerge especially clearly from the German constitutional law on the subject.[17] According to the German Constitution of 1948, the authorising legislation, under which the administrative rule is written, must state clear goals, apply to a clearly-defined subject matter, and set definite limits on the range of possible implementations.[18] The original motivation for these constitutional provisions seem to have stemmed from the Weberian hope that if only statutes can be definite enough, the work of the administrative agencies can be limited to a techni-

[16] Hernando de Soto, "Some Lessons in Democracy—For the U.S." *New York Times*, 1 April 1990, sec. 4, p. 2.

[17] There is a body of American constitutional law, based in the due process clause of the Fifth Amendment, which subjects statutes to invalidation for vagueness. This body of constitutional doctrine, however, is somewhat vaguer than its German counterpart, and, unlike article 80 of the German Constitution, it is not explicitly addressed to the relation between statutes and the administrative rules that fill them out.

[18] The Constitutional Court of the Federal Republic of Germany (Bundesverfassungsgericht, BVerfGE] has struck down several laws because they were too vague to allow any regulation written under their authorisation to meet these requirements. For a general discussion, see (1951) 1 *Entscheidungen des Bundesverfassungsgerichts* 14. Two examples of statutes struck down are the following: a statute establishing a programme of compensation for prisoners of war that left unclear whether it was authorising the issuing of regulations merely setting standards of proof for claims under the act or also establishing who should count as an eligible recipient under the act (BVerfGE 5 [1956] 71); and a statute imposing a value-added tax which left it wholly to the agencies to determine what would be meant by a crucial distinction between "single-stage" and "multiple-stage" enterprises, not even suggesting any criteria in terms of which the appropriateness of a definition could be assessed (BVerfGE 7 [1958] 282, 294; cp. BVerfGE 10 [1959] 251).

cal, instrumental exercise. That may well be how the founders of the Federal Republic of Germany conceived of a well-ordered *Rechtsstaat*. As I have argued elsewhere,[19] and as the pervasiveness of broad delegations to administrative agencies indicates, this hope is a mirage. Agencies are unavoidably involved in settling ends. Yet even under this revised understanding of the situation, these constitutional limits are still crucial. Unless the statutes state clear goals, apply to a defined subject matter, and set definite limits to the range of possible interpretations, the decision-making in the democratically elected legislature will have come to naught. Unless the statutes that the legislature passes give the agencies some definite guidance, any claim that the rule of law is also rule by the people will be ridiculous. I will not pause, here, to battle those post-modernists who hold that no text, legal or otherwise, binds any interpreter. Such views overreact to the failure of narrow conceptions of rationality that stemmed from the logical positivists.[20] The practical question, here, is not whether any legal texts whatsoever can constrain action, but whether the statutes passed by legislators are definite enough to provide a positive basis from which agency deliberation can proceed. Such definiteness is required if the democratic rule of law is to exist at the agency level.

In short, then, for democracy to be possible, two requirements relevant to the administrative agencies must be met. First, the agencies must themselves make rules in a way that accords with the three-fold requirements of the rule of law. The rules must be general, they must provide a predictable basis for citizen action, and they must be generated via a regular and fair process. Secondly, the rules that agencies make must be appropriately constrained and guided by statutes passed by democratically-elected legislatures.

Clearly, however, satisfying these two requirements is still not sufficient for administrative rule-making to be democratic. While, in themselves, they offer no surprises, these two requirements will guide us as we explore ways—possibly new ways—of constraining and conceptualising administrative rule-making so that it may be compatible with the democratic rule of law.

V. WORKING OUT THE DEMOCRATIC RULE OF LAW AT THE ADMINISTRATIVE LEVEL

Any solution to the problem of reconciling democracy with administrative rule-making will have two main components, one more institutional than conceptual, and the other the reverse. The primarily institutional problem is how better to subject administrative rule-making to democratic control and steering. The primarily conceptual problem concerns what relation, precisely, ought to hold

[19] See my "Democracy and Administrative Rationality" (in progress).

[20] For an effective criticism of this sort of post-modernist view, see Martha C. Nussbaum, "Skepticism about Practical Reason in Literature and the Law" (1994) 107 *Harvard Law Review* 714–44.

between statutes and the administrative rules implementing them. These two problems are naturally connected, for a principal function of any institutional arrangements will be to constrain rule-making to comport with the desired relationship, while an important desideratum for any conception of the relationship between statute and administrative rule is that it can be instituted. In the remainder of this essay, I will pursue these two interrelated components by critically examining the views of three recent theorists, Thomas Christiano, Mark Seidenfeld, and Jody Freeman. Christiano suggests a neat, dichotomous solution to the conceptual problem, one that, as I will argue, turns out to be too neat. Seidenfeld, in contrast, is institutionally a quietist. We will learn from why he sees no problem. Freeman puts forward an attractive, "collaborative" model, but one that requires some revision if it is to accommodate actual politics.

Let me start, then, with the conceptual relation between statute and rule, and Christiano's view of it. Of course, we already know, in outline, what relationship we demand between statute and administrative rule. The rule must implement the statute. But how, more precisely, shall we understand this? There are tempting interpretations that are too restrictive, and attractive ones that are too loose. Many theorists have been drawn to the idea that administrative rules must be confined to spelling out efficient means to the end set by legislation. That would be one clear way to envision the idea that the administrative rules merely "implement" the statutes. I have already mentioned my conviction that, contrary to what this suggestion presupposes, agencies cannot avoid setting ends. Now I will pursue this matter in more detail, in order to bring out a crucial ambiguity underlying Christiano's attempt to model the ideal relationship between statute and administrative rule in terms of a sharp dichotomy between end and means.

In his fine book, *The Rule of the Many*,[21] Christiano has championed a conception of democracy oriented around the ideal of equality in joint deliberation. According to Christiano, democracy is the only just mode of government, because only democracy affords citizens an equal opportunity to control the rules under which they live. This equality, he suggests, is not to be measured solely in terms of raw power or influence. Rather, it has two components. Democracy involves a "quantitative" mode of equality, as in the principle of "one person one vote". Yet the vote is only the decisive stage of a broader process of deliberation. In that deliberation, equality is also important; but it is also less easily measured, for deliberation is content-based. It proceeds according to the give-and-take of argument. Accordingly, equality in deliberation must, according to Christiano, be understood "qualitatively", as affording all viewpoints and all arguments equal opportunity to be heard. While this is admittedly not a precise notion, we get the idea. Putting these components of democracy together, we see that what matters is that citizens have equal ability, via argument and deliberation in a fair process open to all, to influence the rules under which they live.

[21] (Boulder, Colo.: Westview, 1996).

It is against this theoretical background that Christiano considers the difficulties posed by the fact that our influence over the rules under which we live is necessarily rather indirect. His solution to this difficulty rests on the claim that "those who choose the aims of a society are the ones who hold decision-making authority".[22] He applies this principle at two separate levels. His first application is to the relationship between voters and legislators. The aim of a just democracy, Christiano argues, ought to be to ensure that elections communicate ends:

> "Once a legislator is elected, he has the duty to represent the aims of the citizens for which he was elected. He no longer has the discretion to change his conception of the appropriate ends".[23]

To make it more likely that parliamentary elections do serve this role, Christiano argues, democracies ought to adopt proportional representation schemes in which party platforms and party discipline play a greater role than in the American government. (He also renews John Stuart Mill's argument that only proportional representation comports with the requirement of quantitative equality.) If parties articulate ends, and if they also exercise considerable disciplinary control over their members once they are elected, then it can be true both that elections endorse ends and that legislators are faithful to them. Legislators must, then, be so circumscribed. "Giving legislators authority to change aims", Christiano writes, "would be an arbitrary infringement on the right of citizens to be equal members of the political community".[24]

When he comes to the stage of agency policy-making, Christiano employs the split between ends and means in the same way. Noting that the activity of the administrative agencies is too complex to be directly controlled in any detail by legislators, Christiano marks the threat to democracy but then sketches how it may be tamed:

> "[This discretion] seems to bring with it the threat that citizens lose control over the government. However, giving discretion to branches of government does not entail lack of significant control. As long as the citizens retain the ability to choose the aims of society and the administration is committed to implementing those aims, the citizens are still in control of what matters for democracy".[25]

Now, at this point, we are entitled to become suspicious. For one apparent implication of Christiano's comments is that whatever it is that the legislators do does *not* matter for democracy. On Christiano's picture, the ends are articulated by political parties. A particular political party's conception of the aims of government are then either endorsed, or not, at the polls. Once this has happened, the suggestion would appear to be, the people have exercised all the control that matters for democracy.

[22] Ibid., p. 216.
[23] Ibid., p. 219.
[24] Ibid., p. 217.
[25] Ibid., p. 239.

But this is absurd. Even under the best-regulated system of proportional representation, popular elections can do no more than select between two or more highly vague political orientations. Perhaps one party is economically conservative but socially progressive, another fiscally more liberal but socially cautious, and so on. One of them is elected. Now the question arises whether to pass a law demanding that public transportation systems be made accessible to the disabled. While these vague orientations may bear differentially upon such an issue, neither speaks to it at all definitively. Either way, this will have to be an issue on which legislators settle ends. Then, when the matter comes to the agencies, they may have to decide what making public transportation accessible to the disabled ought to mean. Should it mean building ramps and buying kneeling buses, or rather providing publicly funded vans?[26] Here, too, the drafters of administrative rules will have to settle ends.

Recall that Christiano's guiding principle is that "those who choose the aims of a society are the ones who hold decision-making authority".[27] This he conjoined with a rhetoric suggesting that all of the decision-making authority could be concentrated in the hands of the citizens. Putting such weight on this principle reveals its notion of "choosing aims" to be fatally ambiguous. If, as I have just suggested, those who choose between mainstreaming the disabled and making special provision for their transportation are choosing aims, then it cannot be the case that all decision-making authority rests with the entire citizenry. If, on the other hand, "choosing aims" is restricted, implicitly, by some such qualifier as "choosing the principal aims, in outline", then this principle, itself, is vague in a troublesome way. Which are the principal aims of a given political society? And what constitutes an "outline" thereof?

Perhaps these difficulties are unnecessary. Perhaps we should simply trust that agencies, either on their own, or in collaboration with interested parties, will arrive at reasonable compromises of competing ends. This is the message of some recent legal theorists who have addressed the problem of administrative discretion.

The idea that we ought to trust the agencies to make reasonable rules has been recently defended in the pages of the *Harvard Law Review* by Mark Seidenfeld.[28] Describing his view as a "civic republican" one, he champions reasonable deliberation by the well-informed as against bargaining and influence-trading among representatives of special interests. We are in a period of reaction against the interest-group theories of politics which were dominant in the 1950s. The drive to find an alternative way of conceiving of politics has helped power the burgeoning interest in deliberative democracy. Seidenfeld combines this fashionable focus on deliberation with a normative suspicion of special inter-

[26] Cf. Robert A. Katzmann, *Institutional Disability: The Saga of Transportation Policy for the Disabled* (Washington, D.C.: Brookings, 1986).

[27] Christiano, *The Rule of the Many*, above n. 21, 216.

[28] Mark Seidenfeld, "A Civic Republican Justification for the Bureaucratic State" (1992) 105 *Harvard Law Review* 1512–76.

ests. As befits a civic republican, his analysis strongly echoes Rousseau. These echoes are particularly clear in the stark way he draws a contrast between private interests and the public good. The aim, he suggests, should be "to focus the debate on a professional understanding of the public interest rather than on accommodation of private interests".[29] The public interest, apparently, does not essentially involve an accommodation of private interests. What "private interests" are, apart from the interests of individuals, is not defined in the article; but it would be consistent with its overall thrust to interpret the notion in a Rousseauvian way as those interests that are not generalisable in the way required of the content of the general will. Building on this sharp division between private interests and the public interest, Seidenfeld constructs an artificial opposition between deliberation and bargaining.[30] Deliberation is oriented toward the public good, whereas bargaining is a matter of the pulling and tugging of private interests.

These contrasts are too sharp. Because the public is essentially—if not exclusively—made up of the individuals in it, the public interest bears an essential relationship to the interests of individuals. Even a social democratic ideal may coherently declare that government must look out for the well-being of *each* individual citizen.[31] Perhaps if there were a workable version of Rousseau's filter that would exclude all interests except those that were, in the appropriate way, generalisable, we could construct a set of "private interests" walled off from the public interest. They would be those that failed the test of generalisability. Yet although Kant's ethics sought to sharpen Rousseau's test, it has, by most accounts, failed at this. While some interests are recognisably sociopathic, we have no general tests for which interests are compatible with the public good and which are not.

Accordingly, it is no surprise that when Seidenfeld turns to defending the administrative agencies as a locus of decision-making, he does not insist on his sharp dichotomy between private interests and the public interest, with sound deliberation being directed only toward the latter. Instead, he describes the agencies as occupying a golden mean. He concludes his article with the following words:

> "Congressional procedures today . . . are unduly influenced by powerful political factions and are not capable of providing sufficient policy coordination to satisfy civic republicanism's mandate of deliberate decision making . . . At the other extreme, courts are too far removed from the values of the polity to satisfy civic republicanism's goal that citizens determine the common good. Administrative agencies, however, fall between the extremes of the politically over-responsive legislature and the under-responsive courts. With proper constraints on bureaucratic decision-making, the agencies' place in government, the professional nature of the agencies' staff, and the

[29] Ibid. at 1554.

[30] E.g., ibid. at 1545.

[31] See Martha C. Nussbaum, "Aristotelian Social Democracy" in R. B. Douglass, G. Mara, and H. Richardson (eds.), *Liberalism and the Good* (New York: Routledge, 1990), pp. 203–52.

> procedures agencies have traditionally used to set policy suggest that the administrative state holds the best promise for achieving the civic republican ideal of inclusive and deliberative lawmaking".[32]

On this picture, then, agency officials are sufficiently removed from the fray of politics to be able to temper the input that they gather from procedures of notice and comment, and the like, with their professional judgement about the public good.

The puzzle that this view leaves us is why Seidenfeld would have us put up with the Congress at all. Since, after all, I have been indulging—as a philosopher is wont—in normative theorising, why shouldn't we consider doing away with the legislative body altogether, if this is correct? And with judicial review of administrative rules, as well? Is it only for purposes of dividing power and providing mutual checks that we should suffer these other branches to continue? I think not. It can only be the refracted influence of the Rousseauvian split between private interests and the public good that pushes Seidenfeld to such a dismissive view of the interest-group bargaining accomplished in Congress. It surely is the case in our system that *monied* interests have too much sway, and for this reason, interest-group lobbying can indeed distort the process's take on the public good. Yet this seems a problem rather with inequality of political resources, of the kind addressed by Christiano's theory, than with interest-group bargaining per se. While we urgently need to address this problem, this is no reason to give up on allowing special interests to try to influence the national legislature. In a large and complex society, public deliberation must depend upon specialised groups first carrying on political discourse and debate in the "informal public sphere", as Habermas would say,[33] and then bringing their causes to the national legislature. If we set aside Rousseau's attempt sharply to separate private interests from the public good, should we not conclude that legislatures ought to be *maximally*—and, of course, fairly—responsive to pressures brought to bear by citizens, whether individually or in groups or firms?

Perhaps what Seidenfeld would want to say is that administrative agencies are to an optimum degree responsive to special interests, *given* that they are working within the constraints imposed by statutes enacted by democratic legislatures. Together with his remarks about the non-responsiveness of courts, this would suggest a neat normative correlation between the degree of generality or abstraction and the degree of responsiveness to democratic pressures. The courts are the least general and the least responsive to popular pressure, the legislature the most general and the most responsive, and the agencies are in the middle on both counts.[34] This position is one to which we might be drawn at the

[32] Seidenfeld, "A Civic Republican Justification for the Bureaucratic State", n. 28 above, at 1576.

[33] See Jürgen Habermas, *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy* (Cambridge, Mass.: MIT Press, 1995).

[34] Compare Mashaw, *Greed, Chaos, and Governance*, n. 13 above, p. 139, where he rationalises this sort of correlation as providing for a balance between democratic accountability, on the one hand, and a flexible justice in individual cases, on the other.

end of the day. I do not mean to argue here that we cannot trust our agency officials to arrive at reasonable policies on the basis of the popular input available to them. But this is just a position, and the argument that we ought so to trust the agencies is missing. More specifically, we lack an argument that so trusting the agencies is compatible with extending the democratic rule of law to the agency level.

In a less complacent vein, Jody Freeman has recently argued for a "collaborative" model of administrative rule-making.[35] She pins her hopes on the types of procedure created by the Negotiated Rulemaking Act of 1990.[36] In cases in which the issue is clear and there is a relatively small group of identifiable interested parties, this Act encourages agencies to involve these so-called "stakeholders" in the *drafting* of administrative rules. The drafting committee is not limited to these previously identified stakeholders: anyone who can make out a significant interest in the matter at issue may ask to participate. The drafting committee is provided with a "facilitator", and stakeholders who are at a serious financial or technical disadvantage may, in certain circumstances, apply for aid. The drafters are meant to come to a consensus, though they are given some leeway in defining what they will take consensus to be. This process of negotiation supplements, rather than supplanting, the stage of notice and comment, for rules that are so drafted still go through the ordinary process before becoming law. To date, negotiated rule-making, or "reg-neg", as it is called, has been most common in environmental regulation. The reason Freeman sees great hope in this process is that she sees it as an opportunity for collaboration as opposed to interest representation. By this she has in mind an open-ended, non-adversarial problem-solving orientation in which there is no great divide between agency "insiders" and petitioning "outsiders".

In light of the appropriate role for interest groups in democratic politics, Freeman's contrast between collaboration and interest representation is less than helpful. As she herself admits, there is a sense in which inviting clearly-identified stakeholders to help draft administrative rules amounts to an *intensification* of interest representation. She even quotes Judge Posner to this effect, in a decision holding that agencies are not bound to promulgate the rules in the form agreed to by such drafting committees. He puts the point in a negative light, writing that taking the agencies so to be bound would be "an abdication of regulatory authority to the regulated, the full burgeoning of the interest-group state, and the final confirmation of the 'capture' theory of administrative regulation".[37] Largely because of her recognition of this, Freeman ends her article on an aporetic note. She sees great promise in the collaborative approach of

[35] Freeman, "Collaborative Governance in the Administrative State" (1997) 45 *University of California Los Angesles Law Review* 1–98. I am grateful to David Dyzenhaus for referring me to this article.

[36] See David M. Pritzker and Deborah S. Dalton (eds.), *Negotiated Rulemaking Sourcebook* (Administrative Conference of the United States, 1995).

[37] *USA Group Loan Services v Riley* 82 F.3d 708, 714 (7th Cir. 1996), quoted by Freeman, "Collaborative Governance", n. 35 above, at 88.

"reg-neg", but she does not see how to purge it of this pesky interest-representation aspect.

I would submit that there is no need for this purgative. As I have just argued, in criticising Seidenfeld, the representation of interests is an essential feature of democracy. The contrast between collaboration and the reigning approach needs to be otherwise conceived. Freeman points out many of the distinctive features of negotiated rule-making that make it collaborative: it is not adversarial in format; it effaces, temporarily, the boundary between those inside the agencies who participate and the other interested parties; all participating are on a par in the attempt to forge consensus, at least at some level; and they must collaborate, because they have a joint task, namely to draft a rule that can gain consensus within the committee. Freeman dislikes the interest-representation model insofar as it involves adversaries who operate with interpretations of what will serve their interests that are fixed in advance. The sort of bargaining that this will produce, as she notes, is epitomised by the idea of splitting the difference. In negotiated rule-making, by contrast, she recognises the fruitfulness of a more open-ended process, in which participants, through their interactions, may well come to reconceive where their interests lie or what solutions might be serve them. This does not mean, however, that the notion of representing interests must be discarded. Rather, since those with a stake in policy questions do have competing interests, we should welcome a process like negotiated rule-making in which those interests can be reinterpreted in creative and potentially conciliatory ways. What we must hope for, at the administrative level and more broadly, is that those who start out on opposite sides of a policy question will sometimes stand ready to revise their ends on the basis of arguments that their opponents offer. If they do, that is a case of what I have elsewhere called "principled compromise".[38]

Freeman's idea that we understand administrative rule-making as a collaborative effort, amended so as to recognise the importance of representing interests, is inherently more democratic than Seidenfeld's suggestion that we simply trust our administrators. The collaborative model also carries to the administrative level a general approach to deliberative democracy that I have set out elsewhere, which analyses democratic deliberation as aimed at forging joint intentions.[39] The notion of arriving at joint commitments about what we shall do, I have argued, is the best way to reconcile two aspects of our democratic ideals that are otherwise at odds. On the one hand, we like to think of democratic debate as truly deliberative, and as governed, when it works well, by a sincere effort to work out what really is in the public interest. On the other hand, we also believe that individuals' mere interests count for *something* in determining where the public good lies. In forging joint intentions through a process

[38] I develop the notion of principled compromise in "Democratic Deliberation about Final Ends" (in progress).

[39] In Henry S. Richardson, "Democratic Intentions", in J. Bohman and W. Rehg (eds.), *Deliberative Democracy* (Cambridge, Mass.: MIT Press, 1997), pp. 349–82.

of democratic deliberation, we figure out what we ought to do in a way that necessarily takes account of the interests of those deliberating. One reason that the negotiated rule-making model is attractive, then, is that it invites those with affected interests to engage in a process with a genuine "we". As I have noted, that is the most distinctive feature of this new process, that it involves its participants in a joint task of framing a policy that all of them can live with. Hence, the negotiated rule-making process is a potentially great improvement over the standard process. To be sure, this case for negotiated rule-making is highly abstract. We lack sufficient experience with the process yet to give an overall verdict on it. A crucial question from which my praise has abstracted is *which* interests are represented. The "we" that is present in these committees may well fail to be adequately representative of the public at large. In particular, while I have made the case that representation of special interests is a necessary feature of democratic politics, it is also important that the point of view of the non-organised public without any special interest at stake be heard. Of course, the negotiated rule-making process is not likely to be worse at giving this part of the public a voice than is the ordinary notice-and-comment process. In both the rule-drafting and the rule-amendment phases, this problem still awaits an institutional solution.

Let me take stock. We saw that conceiving of the democratic rule of law at the administrative level in terms of a sharp split between setting ends and selecting means to them rests on the dangerous fiction that end-setting is over before policy determination is. We then moved to a discussion of an opposite view that, rather than attempting rigidly to constrain the agencies in some such way, simply suggested that we trust them. Yet that approach is, we saw, if not naive, at least unresponsive to our question about how it is that administrative rule-making can be compatible with the democratic rule of law. The collaborative model illustrated by negotiated rule-making, as I will now explain, offers a more promising basis on which to work out a satisfying answer to this question. What we need is an understanding of the agencies' role that recognises that they are inevitably engaged in setting ends and yet constrains this activity in such a way as to keep it democratic.

I said that this question has two levels, an institutional one and a conceptual one. We first discussed the end-means constraint, which is one account of the conceptual relationship that ought to hold between statute and administrative rule, and then shifted to the institutional innovation of negotiated rule-making. My suggestion is that we marry this sort of collaborative institutional model with a different conceptual relation. As my discussion of negotiated rule-making brought out, one of the really promising features of that process is that it will encourage those involved in it to revise their ends on the basis of arguments by other participants so as to arrive at a principled compromise. It is essential to capturing this advantage that the process as a whole involve a joint effort at reformulating ends. The conceptual relation that will keep this reformulation from straying from its proper statutory basis is the relation of specification.

An end is a specification of another if it in effect adds clauses relevantly describing what the end is or where, when, why, how, by what means, by whom, or to whom the action is to be done or the end is to be pursued. "Making mainstream transportation systems accessible to the disabled" is a specification of "making transportation accessible to the disabled". For one end to be a specification of another, it must be the case that every way of satisfying the former is also a way of satisfying the latter. That, according to some, was one problem with the way the Supplemental Security Income programme had been applied in the case of children: *not* every case of supplying benefits to children who had not achieved age-appropriate development was obviously also a case of supplying benefits to disabled children. Accordingly, the relation of specification embodies a real limitation on possible revisions, while also leaving considerable leeway.[40] Neither should be underestimated. The leeway is considerable, for vague aims may be specified many different ways. Further, attempts to specify aims whose vagueness is the product of a stalemate will be especially controversial. Nonetheless, even a vague aim marks out an important range of steps that would *not* count as a way of pursuing it.

It should be clear how specifying ends differs from selecting means to ends. There are two main differences.[41] First, a more specific version of an end does not pick out a causal step in obtaining an end, as does a means. Secondly, the specified end is still an end: that is, it articulates something we pursue for its own sake. A means, by contrast, we do not choose or pursue for its own sake. Together, these differences imply that the effort at specifying public ends is not a matter for predictive experts, but gets at what we fundamentally care about. I have argued that the agencies often are engaged in specifying ends without acknowledging that this is what they are doing. Central to my proposal is the thought that we will not achieve democratic control over the agencies unless we recognise that this is what is going on and tailor our conceptions of public reasoning and the rule of law to accommodate this fact. Nor is it an adequate answer to say that end-setting should be reappropriated by the legislature: in criticising Christiano's view, I have argued that this proposal demands the impossible.

In stating my illustrative cases of specification, I have aimed at brevity. The ends that the agencies should be bound to specify when they make rules, however, will be considerably more complex and intricate. They will already have been somewhat specified by the legislature. Where? Much jurisprudential ink has been spilt on the question of "legislative intent". I have no magic formula to propose for resolving the difficulties of statutory interpretation that arise in

[40] I define the relation of specification more precisely in my *Practical Reasoning about Final Ends* (Cambridge: Cambridge University Press, 1994), sec. 10.

[41] For the purposes of drawing these contrasts, I set aside the intermediate notions of a "constituent means" and of a means that is at the same time an end. Taking these possibilities into account would greatly complicate the exposition of the contrasts without much affecting their point.

cases before the courts. If we are looking for the ends embodied in a particular statute, though, it seems plain that limiting ourselves to explicit statements of purpose in a Bill's preamble will be to cast our net too timidly, whereas allowing administrators to delve for unstated, "underlying" purposes threatens to undermine the constraint that comes with the specification relation. Accordingly, it seems most sensible to take the statute as a whole as articulating what the agency must go on to specify.

A statute is never confined to a statement of ends, whether explicit or implicit. The bulk of a statute will consist in statements of requirement and permission, as well as provisions empowering or disempowering agencies and people in certain respects. The ends embodied in a statute must be specified in light of this whole package. Consider an example from Canadian administrative law.[42] An agency is empowered by statute to promote "equality and fairness" in bargaining between unions and employees. Contemplating the possibility that the enterprise is sold, the statute also requires that the new employer abide by any collective agreement existing at the time of the sale until the employment contract is renegotiated. Within this broad statutory framework, the kind of problematic case that arises will concern, for example, situations in which a company sells off only a portion of its business to a company that employs non-union labour. The agency will need to settle whether the provisions that the statute had contemplated for the sale of the whole enterprise should also apply to such a case. To answer this question in line with my proposal, it is not sufficient simply to attempt to specify "equality and fairness", nor is it sufficient to specify the provision on sales without reference to equality and fairness. Instead, what the agency ought to do is to find an equitable (or egalitarian?) and fair way of specifying the provision about sales. Alternatively put, the agency ought to spell out how it plans to understand the end (or ends) of equality and fairness in the context of partial sales.

Here, then, is the picture: members of the public press certain general causes, such as making transportation systems accessible to the disabled. When the legislature acts on such an issue, it generally will lay down some parameters. Even if it does not settle matters such as the mainstreaming issue, the statute in question will always include some specifications, such as which agency is to carry out the work or how any costs incurred are to be paid for. However the end of providing transportation to the disabled is specified in the statute, those drafting any administrative rule implementing the statute must work within it. The task of a drafting committee, whether internal or including stakeholders, is to work out a fuller specification of doing *that*, whatever it is. Each of the policy-making stages, legislative and administrative, requires democratic input, for each is a non-technical exercise in setting public ends. In this respect, they are the same in kind. Yet there is a lot of sense in dividing our labour so that policies unfold in

[42] I owe this example, and all my knowledge about its particulars, to David Dyzenhaus. It does not matter, for these illustrative purposes, whether the details are correct.

stages in this way. We cannot all get together to work out our policies in detail, all at one shot. In order for us to work them out democratically in any detail, though, the more abstract, initial decisions that are made by legislatures must be honoured by those at the administrative level who gather to work out the specifics.

In a way, this proposal highlights the insufficiency of the rule of law for democracy. By insisting on separate democratic input at the administrative level, it invites one to reproduce my general argument for the insufficiency of the rule of law for democracy. Just as the rule of law can characterise a constitutional monarchy, so, too, can it characterise a government—like Peru's some time back, perhaps—that has a democratically elected legislature but cedes all control over administrative rule-making to a tiny oligarchy. Yet my picture of democratic policy-making also gives great importance to the rule of law. As I have argued, carrying the rule of law to the administrative level is a prerequisite of democracy. The administratively generated rules under which people live must be general, predictable, and framed by a regular process. The demand that these rules count as specifications of statutes adds an essential and obvious layer of predictability and regularity. It brings out a corollary of the rule of law that I have not yet stated. If the law is to be predictable, it must not be rife with conflict. It will be no good for the citizens if the legislature generates one set of clear and general rules about sulfur dioxide emissions, say, and the administrative agencies generate another, completely unrelated set. For the rule of law to be possible, the law must have more unity than that.

Obviously, the way to avoid this chaos is for the administrative rules to be subordinated to the statutory ones, in just the way that the specification relation sets out. Accordingly, working out how to extend the rule of law to the administrative level is not only necessary to maintaining a real democracy—it is also fruitful in conceiving what a real democracy should be.

16

Private Parties, Public Functions and the New Administrative Law

JODY FREEMAN[1]

I. INTRODUCTION

Reconciling administrative power with democracy has long pre-occupied American administrative law scholars. Agencies inhabit a precarious position in the American separation of powers regime. They are the "headless fourth branch,"[2] for which scant provision is made in a Constitution that divides powers among the Congress, a unitary executive and the judiciary.[3] Although not directly accountable to the electorate, agencies wield enormous discretionary power in the implementation of their delegated mandates; even the most specific statutes leave considerable room for interpretation and discretionary judgments. Because the American democratic system requires that the exercise of governmental authority be accountable to the electorate,[4] administrative law

[1] Acting Professor of Law, UCLA School of Law. Email: Freeman@law.ucla.edu. I am indebted to many colleagues for their insightful comments on earlier drafts: Rick Abel, Michael Asimow, David Dolinko, David Dyzenhaus, Stephen Gardbaum, Mitu Gulati, Joel Handler, Ken Karst, Lisa Krakow, Gillian Lester, Dan Lowenstein, Tim Malloy, Steve Munzer, Mark Seidenfeld, Seana Shiffrin, David Sklansky, Richard Steinberg, Nancy Staudt, and Jonathan Zasloff. Responsibility for errors and omissions is mine alone. I thank Marc Luesebrink for excellent research assistance.

[2] In a 1937 report, a Presidential Commission referred to independent agencies as a "headless fourth branch of government, a haphazard deposit of irresponsible and uncoordinated powers." See President's Commission on Administrative Management, *Report Of The Committee With Studies Of Administrative Management In The Federal Government* (1937), at 40. Thanks to Jonathan Weinberg for tracing this reference to the Brownlow Report and sharing it with subscribers to the administrative law list.

[3] The Constitution does not mention administrative agencies *per se*. Authority for delegating power to them is rooted principally in the Take Care and Necessary and Proper Clauses. Art. II, § 3 requires that the president, "take care that the laws are faithfully executed." Art. I, § 8 empowers congress to "make all Laws which shall be necessary and proper for carrying into Execution the foregoing powers." Art. II, § 2 provides for the appointment of "Officers of the United States." Still, as one scholars notes, "the idea that administrators make law, decide legal claims, or operate outside of the executive branch has troubled legal theorists since the federal government's first forays into civilian regulatory activity." See Jerry L. Mashaw, *Greed, Chaos and Governance* (New Haven: Yale University Press, 1997), at 107.

[4] While this is not the place for an exegesis on democratic theory, some definitions are in order. My use of the term democracy refers to the core requirement at the heart of the uniquely American system: accountability to the electorate. The bicameral, presidential, two-party American system is, of course, only one type of democracy. Numerous political arrangements that might be viewed

has largely organized itself around the need to provide accountability indirectly through various mechanisms designed to discipline agency behavior, including legislative and executive oversight and judicial review.[5] Nonetheless, uncontrolled agency discretion remains a constant threat to the legitimacy of the administrative state.[6] It represents what might be called the traditional democracy problem in administrative law.[7]

as democratic may also have additional features that distinguish them from one another. The shape of any democracy turns on such things as the structure of the party system, whether it is a parliamentary or presidential system, the level of economic development, and the social structure. For most political theorists, a regime that fails to satisfy basic social, economic and political pre-conditions cannot claim to be democratic. Those pre-conditions are hotly contested, however. Dahl argues that the term democracy should be used only to describe the ideal of "a political system one of the characteristics of which is the quality of being completely or almost completely responsive to all its citizens." See Robert A. Dahl, *Polyarchy* (New Haven: Yale University Press, 1971). When administrative law scholars discuss the "problem of agency discretion" or the "absence of administrative legitimacy" or refer to agencies as a threat to "democracy," they usually do so without specifying the theory of democracy to which they subscribe. Instead, they are usually referring to the need for accountability. Direct or indirect accountability enables the electorate, through voting or similar means, to punish or reward their elected representatives for agency actions of which they disapprove. Because agency officials are not themselves elected, they must be subject to oversight by both elected officials, who can themselves be held to account for agency actions, and the judiciary, which ensures that administrative decisions comply with the rule of law.

[5] See Kathleen Bawn, "Choosing Strategies to Control Bureaucracy: Statutory Constraints, Oversight, and the Committee System" (1997) 13 *Journal of Law, Economics, and Organization* 101. See also Clinton Executive Order No. 12,886 requiring cost-benefit analysis for "major rules" and annual regulatory impact analysis; The Unfunded Mandates Reform Act of 1995, Pub. L. No. 104–4, reprinted in 1996 *U.S.C.C.A.N.* (109 Stat.); The Small Business Regulatory Enforcement and Fairness Act, 5 *U.S.C.A.* § 801 (West Supp. 1996). Judicial review is meant to guarantee the legality of agency decision making. I use the word legality to mean two things: fidelity to legal procedure and compliance with substantive norms of rationality. The distinction between procedural and substantive regularity is admittedly somewhat artificial in American administrative law. Often courts will treat compliance with legal procedure as sufficient indication of substantive rationality. Nonetheless, judicial review under the American Administrative Procedure Act entitles courts to invalidate agency action that is substantively "arbitrary and capricious," regardless of procedural compliance. See Administrative Procedure Act 5 U.S.C. § 706 (1994).

[6] Sometimes scholars use the word legitimacy interchangeably with the word accountability. In the administrative law context, I understand legitimacy to mean the *acceptability* of administrative decisions to the public, which might in part derive from accountability, but which might derive from other things as well. For more on the distinction, see note 14 below.

[7] Administrative law scholars have traditionally viewed administrative discretion as the greatest problem of the field. See Kenneth Culp Davis, *Discretionary Justice: A Preliminary Inquiry* (Baton Rouge: Louisiana State University Press, 1969) (". . . the central inquiry is what can be done that is not now done to minimize injustice from exercise of discretionary power." *Id.* at 1; ". . . discretionary power is a necessary governmental tool but excessive discretionary power is dangerous and harmful." *Id.* at 27; "What is obviously needed is balance—discretionary power which is neither excessive nor inadequate.... In the United States today, all levels of government—federal, state, and local—lack that balance. Our governmental systems are saturated with excessive discretionary power which needs to be confined, structured, and checked." *Id.* at 216.).

For an example of an article wrestling with the problem of adequately controlling agency discretion in the rulemaking context, see Henry Richardson, *Administrative Policy Making: Rule of Law or Bureaucracy*, in this volume. Richardson's article illustrates the difficulties of attempting to provide administrative legitimacy through discretion-constraining legal tests (such as his proposed specification test). It exemplifies how the traditional administrative law concern over controlling agency discretion tends to degenerate into debates over interpretive methodology (how to determine Richardson's "End 1") and standards of review (what deference is owed to agency decisions). While

In this essay, I argue that persistent attempts to solve this democracy problem have obscured a phenomenon with perhaps even greater implications for democracy: the role of non-government actors in the exercise of administrative authority. An *exclusive* focus on agency discretion prevents us from appreciating the extent of private participation in governance. Although legal scholars have long acknowledged some forms of private engagement with public agencies, such as lobbying and litigation, non-government actors remain marginal in a field dominated by agency action. And yet, non-government actors are involved in all stages of the regulatory and administrative process, sometimes assuming or sharing roles that we think either are, or ought to be, reserved for public actors. To the extent that it has been discussed in legal scholarship, the participation of private actors in the administrative process is framed either as a delegation issue or an illustration of public choice theory.[8] Models of the administrative process, whether pluralist, civic republican or "expertocratic" continue to focus primarily on agencies as decision makers.[9]

In fact, many private actors participate in governance in ways that are rarely recognized by the public, acknowledged by politicians or carefully analyzed by legal scholars.[10] The contributions of private individuals, private firms, financial institutions, public interest organizations, domestic and international standard-setting bodies, professional associations, labor unions, business networks, advisory boards, expert panels, self-regulating organizations and non-profit

undeniably one of the "persistently intriguing puzzles" in American Administrative Law, the nature of judicial review of agency action is not the only puzzle worth solving. See Cynthia Farina, "Statutory Interpretation and the Balance of Power in the Administrative State" (1989) 89 *Colum. L. Rev.* 452. Indeed, a pre-occupation with the relationship between agencies and courts can marginalize other important relationships, especially those between agencies and private actors responsible for designing and implementing public policy.

[8] On the constitutionality of private delegation, see, e.g., Harold J. Krent, "Legal Theory: Fragmenting the Unitary Executive: Congressional Delegations of Administrative Authority Outside the Federal Government" (1990) 85 *Nw. U. L. Rev.* 62. For foundational works on interest group pressure, see George Stigler, *The Citizen and the State: Essays on Regulation* (Chicago: University of Chicago Press, 1975); Mancur Olson, *The Logic of Collective Action: Public Goods and the Theory of Groups* (Cambridge, Mass.: Harvard University Press, 1965).

[9] See Jim Rossi, "Participation Run Amok: The Costs of Mass Participation for Deliberative Agency Decisionmaking" (1997) 92 *Nw. U. L. Rev.* 173 (explaining how public participation is understood within different theoretical models of agency decision making).

[10] I do not, however, wish to overstate the extent to which agencies are thought to act alone. Administrative law scholars do recognize, for example, that private citizens play a role in the enforcement of statutes through citizen suit provisions and that private parties also participate extensively in the informal notice and comment process. Scholars acknowledge as well, the considerable informal influence exercised over Congress and administrative officials by regulated interests and other well-financed groups. See William Greider, *Who Will Tell the People: The Betrayal of American Democracy* (New York: Simon & Schuster, 1992) (on the role of think tanks in providing statistical information to members of Congress). The dangers of pluralism as a mode of administrative decision making are well understood in administrative law. See Thomas McGarity, *Reinventing Rationality: The Role of Regulatory Analysis in the Federal Bureaucracy* (Cambridge: Cambridge University Press, 1991). Still, public participation is treated in legal scholarship as a way to influence agency decision making. The exertion of influence is very different from direct responsibility for policy making or implementation. The activity of interest groups and private actors is never considered in theory to itself constitute government action.

groups belie administrative law's pre-occupation with agency discretion.[11] Private individuals serve on influential government boards; "expert" private committees exercise important powers of accreditation; private producer groups formulate regulations and set prices that bind dissenters; private groups may directly negotiate regulations together with other interested parties and the agency; non-profit and for-profit organizations contract to provide a variety of government services and perform public functions ranging from garbage collection to prison operation; individuals and organizations act as private attorneys general in prosecuting statutory violations; trade associations generate and enforce industrial codes that may become de facto regulatory standards; private standard setting organizations generate health and safety standards that agencies automatically adopt. Contemporary governance might be best described then, as a regime of "mixed administration"[12] in which private and public actors share responsibility for both regulation and service provision. In Part II, I provide illustrations of this phenomenon.

The agency emphasis in administrative law makes it difficult to imagine the appropriate legal and institutional response to the essentially shared nature of governance.[13] It threatens to misdirect political and scholarly energy toward imposing ever more marginally effective controls on agencies when in fact agencies are part of a richer institutional environment of public and private activity. Many scholars believe that private actors exacerbate the lack of accountability that makes the exercise of agency discretion so problematic. Like administra-

[11] In legal theory, this degree of private delegation raises concerns about both the accountability of private groups and the threat delegation poses to separation of powers principles, such as the Article I interest in Congress making the laws and the Article II interest in a unitary executive. See Krent, "Legal Theory: Fragmenting the Unitary Executive: Congressional Delegations of Administrative Authority Outside the Federal Government," note 8 above. In contrast to legal scholars, implementation theorists and federalism scholars outside of law acknowledge the potentially positive roles played by private actors in governance. See John T. Scholz, "Federal vs. State Enforcement: Does it Matter?" *in Power Divided: Essays on the Theory and Practice of Federalism*, Third Berkeley Seminar On Federalism (Harry N. Scheiber ed., Berkeley, CA: Institute of Governmental Studies, U.C. Berkeley, 1989).

[12] For the term "mixed administration," see Mark Aronson, "A Public Lawyer's Response to Privatization and Outsourcing" in *The Province of Administrative Law* (Michael Taggart ed., Oxford, UK: Hart Publishing, 1997).

[13] Ibid., at 52, defining "shared governance" as a regime in which private actors and government play shared regulatory roles. See also, Steven Rathgeb Smith and Michael Lipsky, *Non-Profits for Hire: The Welfare State in the Age of Contracting* (Cambridge, MA: Harvard University Press, 1993). The terms public and private are a useful, though limited, short hand for describing different kinds of regulatory actors. Because they convey the basic difference between government and non-government actors, I continue to use them throughout the essay. Their major failing, however, is that they conjure up rather cardboard cut-out images of government *versus* private actors. They are not very useful for conveying the extent to which different kinds of actors depend upon one another to such an extent that the lines between them blur, or the extent to which different actors overstep the bounds of their definitional category and perform duties that are not formally allocated to them. The regulatory arrangements described in Part II are meant to illustrate this phenomenon. Even traditional command and control measures rely to significant extent on the cooperation of regulated entities. Is command and control regulation properly described as "government" regulation? If an agency exercises its enforcement discretion favorably for self-regulated firms, is it appropriate to describe the system as a voluntary system of "private" regulation?

tors, private actors are unelected. Unlike agencies, however, they are not generally *expected* to serve the public interest, nor are they subject to institutional norms of professionalism and public service that might militate against the pursuit of mere self-interest or capitulation to narrow private interests.

Moreover, private actors remain relatively insulated from legislative, executive and judicial oversight. To the extent that private actors increasingly perform traditionally public functions unfettered by the scrutiny that normally accompanies the exercise of public power, private participation may indeed raise accountability concerns that dwarf the problem of unchecked agency discretion. In this view, private actors do not raise a *new* democracy problem; they simply make the traditional one even worse because they are considerably more unaccountable than agencies. In addition, private actors may threaten other public law values that are arguably as important as accountability. Their participation in governance may undermine features of decision making that administrative law demands of public actors, such as openness, fairness, participation, consistency, rationality and impartiality.

Concern about how private actors compromise public law values has only intensified in an era of widespread privatization and contracting out of government functions. To defend against the threats posed by increased private activity some scholars propose that we extend to private actors the oversight mechanisms and procedural controls that apply to agencies, effectively treating them as if they were "public." I discuss this proposition in Part III. I argue that we ought to resist the impulse to constrain private actors as we would agencies in the absence of a careful consideration of the advantages they offer and the threats they pose. Importantly, private actors are often *already* constrained by alternative accountability mechanisms that go largely unrecognized in administrative law. A private decision maker's internal procedural rules, market pressures, informal norms of compliance, third party oversight and the background threat of agency enforcement might hold private actors to account for their performance, even in what seem to be voluntary, self-regulatory systems. Although these forms of accountability may not satisfy the traditional administrative law demand for accountability to an elected body, they nonetheless may play an important role in *legitimizing*, or rendering acceptable, a particular regulatory regime.[14] Before we can assess whether alternative accountability mechanisms

[14] I use the term *legitimacy* to convey the notion of the *acceptability* of administrative decisions to the public. I do not treat it as a "core" value of democracy as I do accountability, although some might argue that democracy requires legitimacy or depends upon it. For purposes of the discussion here, decisions are legitimate when the public accepts them without having to be coerced. Legitimacy might derive from the presence of accountability mechanisms. It might also derive from an agency's fidelity to legal procedures and compliance with norms of substantive rationality (the combination of which might be called *legality)*. Although accountability and legality are crucial sources of legitimacy, they are not the *only* sources of legitimacy. A decision might be viewed as acceptable to the public because it appeals to them as simply "right" or because it is the product of a particularly respected decision maker or because it is a technically optimal solution to a regulatory problem. A decision might be *legitimate*, therefore, without the presence of an accountability mechanism, but it is highly unlikely.

might be appropriate substitutes for, or complements to, traditional forms of oversight we must, at a minimum, recognize that they exist.

A deeper understanding of the private role in governance will help clarify both the dangers that private parties represent and the accountability mechanisms, both traditional and non-traditional, with which administrative law might respond. Viewing governance as a shared enterprise allows us to separate the mechanisms that *produce* accountability from the public or private *nature* of the decision maker. This in turn helps to cast private parties in a more realistic and balanced light. Private actors are not just rational, self-interested rent-seekers that exacerbate the traditional democracy problem in administrative law; they are also *regulatory resources* capable of contributing to the efficacy and legitimacy of administration.

I conclude in Part IV by suggesting an administrative law agenda that includes more in-depth study of public-private interdependence in administration. I speculate about why the agency focus remains so entrenched in administrative law despite the pervasive role of private actors. I underscore the need to develop a theory of the state's role in a shared governance regime and the need to develop as well an accompanying theory of judicial review. Finally, I introduce the notion of *regulatory regimes* to replace the current conceptualization of "public" and "private" roles in governance.

II. ILLUSTRATIONS OF PUBLIC-PRIVATE ARRANGEMENTS

Providing examples of private participation in governance is an important first step toward understanding the range and complexity of interdependent relationships to which administrative law must respond. In this section, I briefly describe a variety of public-private arrangements that exemplify the concept of "mixed" administration in which the performance of "public" functions or the exercise of regulatory responsibility is shared. The illustrations below are meant, however, to be suggestive rather than comprehensive. In addition to identifying a variety of public-private arrangements, I also suggest why they raise accountability problems and why the private role tends to provoke the most concern.

The illustrations are preliminary in part due to a dearth of empirical work. While the provision of public services has long been a shared venture that relies on public financing of private providers, only recently have private actors moved into more traditionally public functions such as incarceration. Widespread contracting out of this and other public functions raises important questions about the dividing line between the public and private roles in governance and the appropriate mechanisms for ensuring accountability.

Legal scholars have yet to examine closely the implications of public–private cooperation in standard-setting, implementation and enforcement. For example, mandated self-regulation, an arrangement in which stakeholders design and

implement a regulatory scheme under agency supervision, has attracted little empirical study.[15] So too, with voluntary self-regulation by private firms within a traditional command and control system.[16] Agency sponsored initiatives such as regulatory negotiation, which are aimed at involving private parties more directly in the design of rules and standards, have by now generated a substantial literature, but these studies reflect little on the larger implications of shared regulatory roles or joint responsibility.[17] Beyond the familiar literature on agency capture, one struggles to find careful analyses of how traditional command and control regulation depends so heavily upon private self-reporting, negotiation and industry cooperation in enforcement.

The examples of private participation in regulation (as opposed to service provision) are drawn primarily from my own fields, environmental and occupational health and safety law, but other examples arise in virtually every regulatory context, including securities regulation, agricultural marketing, food and drug regulation and professional licensing. Moreover, while my illustrations are mostly federal or state, one can find public-private arrangements functioning at every level of government. Indeed the blending of public and private actors is likely to be especially complex at the state and local level. Clearly, all of these arrangements and relationships, whether longstanding or new, cry out for empirical research.

A. Privatization/Contracting Out

The most common example of public-private cooperation in governance takes the form of agencies contracting with private non-profit and for profit firms to provide social services or perform public functions.[18] Widespread contracting

[15] An exception is Joseph V. Rees, *Reforming the Workplace: A Study of Self-Regulation in Occupational Safety* (Philadelphia: University of Pennsylvania Press, 1988). There are also numerous articles on mandatory self-regulation in the securities industry. See Gabriel S. Marizadeh, "Self-Regulation of Investment Companies and Advisers: A Proven Solution to a Contemporary Problem" (1997) 16 *Annual Review of Banking* 451.

[16] Most accounts of self-regulation in environmental law have come from political and social scientists. See Rees, *Reforming the Workplace: A Study of Self-Regulation in Occupational Safety, supra,* note 15; see also Joseph V. Rees, *Hostages of Each Other: Transformation of Nuclear Safety Since Three Mile Island* (Chicago, IL: University of Chicago Press, 1994); Neil Gunningham and Peter Grabosky, *Smart Regulation* (Oxford: Clarendon Press, 1998) (see chapter 4 in particular for an analysis of the shortcomings of self-regulation, including lack of transparency and independent auditing, concern that performance is not being evaluated, absence of real penalties for recalcitrants); Ian Ayres and John Braithwaite, *Responsive Regulation: Transcending the Deregulation Debate* (New York : Oxford University Press, 1992).

[17] See Cornelius M. Kerwin and Laura I. Langbein, *An Evaluation of Negotiated Rulemaking at the Environmental Protection Agency: Phase I* (1995) (report prepared for the Administrative Conference of the United States); Cary Coglianese, "Assessing Consensus: The Promise and Performance of Negotiated Rulemaking" (1997) 46 *Duke L.J.* 1255.

[18] On the important role played by non-profit groups in providing publicly financed services, and the larger phenomenon of "third party government" see Lester M. Salamon, *Partners in Public Service* (Baltimore and London: The Johns Hopkins University Press, 1995); Burton A. Weisbrod, *The Nonprofit Economy* (Cambridge: Harvard University Press, 1988).

out , together with privatization in the form of selling state assets, is at the heart of the public sector reform movement in Britain and other liberal democracies. In the United States, many social services have long been funded by government agencies but provided by non-government actors.[19] Other services, such as prison operation, have been traditionally performed by government and only recently relinquished to private actors.[20]

Privatization is an imprecise term;[21] scholars use it to describe very different phenomena, ranging from the sale of state-owned assets to de-regulation to contracting out the provision of goods and services.[22] The term contracting out is probably more accurate in most cases of ostensible "privatization" because it better captures the state's residual capacity to exercise supervisory control. Consider the private management of prisons. While the literature on private prison operation refers to the growth of private prison management as "privatization" I refer to it instead as contracting out because private prisons operate

[19] For numerous examples, see Salaman, *Partners in Public Service*, note 18 above; Smith & Lipsky, *Non-Profits for Hire,* note 13 above; Handler, *Down from Bureaucracy: The Ambiguity of Privatization and Empowerment,* (Princeton, N.J.: Princeton University Press, 1996). See also GAO Report to Subcommittee on Human Resources, Committee on Government Reform and Oversight, House of Representatives, "Social Service Privatization" (documenting expanded privatization of social services and analyzing implications for accountability) USGAO, October, 1997.

[20] See Laura Suzanne Farris, "Private Jails in Oklahoma: An Unconstitutional Delegation of Legislative Authority" (1998) 33 *Tulsa L. J.* 959 (noting the trend toward "privatization" and stating that one hundred and twenty private jail and prison operations are operating in twenty-seven states). Although contracting out prison *management* is different from contracting out the provision of less significant services within a publicly run prison (such as meal service or medical treatment), the difference is one of degree, not kind.

[21] I reserve the term privatization for those cases in which the state sells assets, although even these sales may not amount to the absolute relinquishing of control. Even in an ostensibly "privatized" regime, government usually retains important powers. For example, agencies may still influence industrial policy through licensing or reserving shares in newly private corporations, which gives the government veto power over major decisions and enables it to block undesired takeovers. See C. Graham and T. Prosser, "'Rolling Back the Frontiers'? The Privatisation of State Enterprises" in *A Reader on Administrative Law* (D.J. Galligan, ed., Oxford University Press, 1996).

[22] Economists and policy makers usually justify privatization on the theory that private control will provide efficiency gains. But cf. John D. Donahue, *The Privatization Decision* (USA: Basic Books, 1989) at 35 (arguing that the presence of competition is more important than the public or private nature of the decision maker). Proponents of privatization also argue that the capital market and the company meeting will provide more direct accountability than public agencies. Privatization does not guarantee accountability, however. Often it enables secrecy and impedes the flow of information. It may also insulate private actors from liability for constitutional violations. Shareholders do not call companies to account very easily. In some cases, private ownership has reduced access to information that was more easily available under public ownership. See Graham and Prosser, n. 21 above.

Critics wonder whether contracting out will lead to the pursuit of efficiency at the expense of the substantive goals of the regulatory program. Some government functions, including regulating pollution, providing an adequate housing stock, eliminating discrimination, ensuring quality education or providing for the poor, seem incompatible with efficiency as an end in itself and unlikely to be effectively privatized. Indeed, government entrance into each of these areas (assuming government was not already present) was justified in part by market failure and cannot be adequately accounted for in a public choice framework. The prospect of widespread privatization forces us to confront the potential conflict between efficiency and other values, such as equality. Critics of privatization argue that in a largely privatized system, market values will be systematically elevated above others.

in relationship to government agencies, courts and legislatures, which might exercise supervisory control through revocable delegations, contract and licensing.[23]

Advocates of private prison management claim, as do most proponents of privatization in other contexts, that private management offers numerous advantages over government management: prisons will be built faster and cheaper, they will be run more efficiently and services will improve. To date, the legal commentary on "private" prisons focuses not on the quality of service but on two more fundamental issues: the constitutionality of the delegation of public power to private actors and whether, if such delegation is constitutional, a private, for-profit company operating a prison should be treated as a public actor bound by constitutional constraints.[24] The delegation question turns on the specific structure of the relationship between government and private actors. If the government delegates only *management* of the prison but retains rulemaking and adjudicative authority for itself, the delegation will likely survive constitutional scrutiny under existing case law.[25] To ensure a delegation's constitutionality then, a state legislature might reserve to an agency the power to accept, reject or modify any proposed rules that would affect the prisoners. It might also provide for judicial review of private adjudicative decisions.

In most cases, public agencies engaged in contracting out believe they are surrendering only policy *implementation* while retaining authority over policy *making*. The distinction is tenuous at best, however. Private prison guards exercise discretion that affect prisoners' most fundamental liberty interests (over meals, showers, exercise time, cell conditions, transportation, work assignments, visitation). Prison officials judge when infractions occur, impose punishments and make recommendations to parole boards.[26] The distinction between day to day management (a private function) and ultimate rule making and adjudicative power (a public function) may not be sustainable. Even where an agency retains the authority to accept or reject rules proposed by the private provider, the provider *interprets and put into operation* those rules, giving them their practical meaning and blurring the line between the policy making and implementing functions. Authority over day to day operation confers upon the private manager a "governmental" power to both legislate and adjudicate.

One can see how the private prison example illustrates the threat of unaccountability associated with private participation in governance. Unlike other functions with a history of private sector involvement, the incarceration

[23] See Ira P. Robbins, "The Impact of the Delegation Doctrine on Prison Privatization" (1988) 35 *UCLA L. Rev*. 911.

[24] Ibid., see also Joseph E. Field, "Making Prisons Private: An Improper Delegation of a Governmental Power" (1987) 15 *Hofstra L. Rev*. 649.

[25] Robbins, "The Impact of the Delegation Doctrine on Prison Privatization," note 23 above, at 930–34.

[26] Field, "Making Prisons Private: An Improper Delegation of a Governmental Power," note 24 above, at 659.

function has been described as "intrinsically governmental in nature."[27] Contracting out a traditionally governmental function creates significant accountability problems because the private provider is one step further removed from direct accountability to the electorate. Private performance of a government function can also create conflicts of interest between private and public goals. This is especially obvious in the case of prisons, where the private interest in maximizing profits may conflict with the public interest in sound correctional policies. Private managers may choose to cut costs by reducing staff or hiring under- qualified guards, which compromises the safety of both prisoners and staff alike. Private adjudication of infractions and imposition of penalties in the prison context implicates prisoners' constitutional rights.

In response to the perceived dangers of private prison management, most scholars would insist that private prisons submit to oversight by agencies, courts and the legislature, and that they observe constitutional limits—the traditional constraints typically imposed on agencies performing the same functions. However, the mere existence of agency oversight and observance of traditional procedural requirements may be insufficient to protect against the myriad dangers of the regime. Such considerations drive some scholars to simply advocate against contracting out the incarceration function, despite the potential efficiencies of private management.

Although the prison example is particularly provocative because it implicates liberty interests, contracting out more innocuous public functions has the potential to create similar disquiet over the absence of accountability or the loss of other public law values such as participation and fairness. Such concerns arise with almost every example of contracting out, including those that do not implicate liberty interests or that less obviously confer on private actors a policy *making* function. In a typical contractual regime involving public services, the agency and private provider negotiate the terms of the contract and the agency acting as consumer may then seek either to enforce the terms of the contract or terminate it. The contract between agency and private provider is meant to specify the terms under which the private party will implement the agency's policy decisions.[28] A contractual system relies on judicial enforcement of the private law of contract at the behest of the *supervising agency* rather than judicial enforcement of administrative law principles at the behest of *private citizens*. As a result, a contractual regime might undermine public participation in decision making and impede public access to relief for injuries suffered by the intended beneficiaries of the contract. The private contractor may provide poor service, injure consumers or engage in anti-competitive behavior, for example, all with little fear of reprisal.

[27] Robbins, "The Impact of the Delegation Doctrine on Prison Privatization," note 23 above, at 934.

[28] This is reminiscent of the original delegation problem between Congress and agencies and the unsuccessful attempt to maintain a clear dividing line between policy choice and implementation.

To render this more concrete, I rely on Aronson's compelling example of a patient facing eviction from a private provider's government funded nursing home in breach of the terms of the agency-provider contract.[29] Privity of contract may prevent the patient from suing to enforce the terms of the agreement or from recovering for damages for the breach. The consumer may lack the incentive or adequate resources to sue on her own, even if she could recover at common law. Absent vicarious liability, the contracting agency may have no motivation to supervise the provider's compliance with contractual terms.[30] Moreover, the terms of such contracts may not be well publicized, making it more difficult for consumers to enforce them.

A contractual regime necessarily depends on public-private partnership. In the examples above, government actors maintain significant oversight roles but private actors perform key "public" functions. When government contracts out its coercive regulatory authority, however, we might worry more about abdication of responsibility than when it contracts out mere service provision. After all, the traditional explanation of why government provides goods and services is market failure: either the market has no incentive to provide them or the market for such goods would be imperfect (a natural monopoly for example). Contracting out services such as refuse collection may strike us as entirely appropriate if the cost of government provision exceeds the inefficiencies of market imperfections.[31] Moreover, we worry more about the possibility of agencies abdicating their policy *making* functions than we do their contracting out mere *implementation*. Because the distinction between policy making and policy implementation is dubious, however, and because even the private provision of services may compromise things we care about, such as accountability, procedural fairness and participation, I consider both the service and regulatory contexts to raise important questions about private participation in governance.

B. Shared Standard-Setting

The delegation of standard-setting authority is importantly different from contracting out service provision because it explicitly allows the formal delegation of policy *making* in addition to implementation. Notwithstanding the indistinct line between these two functions, commentators tend to think that even greater accountability problems arise when the initial responsibility to establish standards falls to non-government parties. Delegation of standard setting to private parties is problematic in the United States because of the non-delegation

[29] Aronson, "A Public Lawyer's Response to Privatization and Outsourcing," note 12 above, at 65.

[30] Ibid., at 54–55, offering the example of a pensioner having difficulty recovering for damage to his letterbox caused by a private contractor delivering mail for Australia post.

[31] Thanks to Mark Seidenfeld for making this point in his especially helpful comments on the special dangers of the private exercise of coercive power.

doctrine.[32] Nonetheless, both the Supreme Court and many state courts have approved numerous delegations to private actors to set standards in circumstances in which little oversight and few procedural checks encumber private discretion.[33] For example, private producer groups such as dairy farmers and handlers, as well as wheat and tobacco growers, set prices that bind dissenting members of the industry. Producers of other agricultural commodities not only set prices but also establish quotas and determine unfair labor practices.[34]

Moreover, agencies may *indirectly* and less visibly delegate standard-setting authority by adopting privately-generated standards after a cursory notice and comment process. There is a long tradition of agency incorporation of privately established health, safety and product standards.[35] Numerous non-governmental organizations set a variety of standards that very directly affect public health and safety. Such organizations include professional associations (such as the Society for Automotive Engineers and the American Society of Mechanical Engineers), non-profit groups (such as the American Society for Testing and Materials), membership organizations comprised of manufacturers, professionals and government officials (such as the National Fire Protection Association), and trade associations (such as the American Petroleum Institute). These private actors generate thousands of industrial codes (building codes, plumbing codes, fire codes) and product standards that govern the design, material, processing, safety and other characteristics of products.

Although nearly invisible to the public, these codes and standards have enormous economic and social influence. Many of them are incorporated by reference by agencies such as the Food and Drug Administration, the Nuclear Regulatory commission, the Federal Aviation Administration, and the Occupational Safety and Health Administration (OSHA). They are also widely adopted by state and local governments.[36] In some cases, legislation *directs* agency officials to adopt such standards, as with OSHA in its early years.[37] The

[32] See *A.L.A. Schechter Poultry Corp. v. United States*, 295 U.S. 495 (1935); *Carter v. Carter Coal Co.*, 298 U.S. 238, 311 (1936); *Panama Refining Co. v. Ryan, 293 U.S.* 388 (1935).

[33] See Krent, "Legal Theory: Fragmenting the Unitary Executive: Congressional Delegations of Administrative Authority Outside the Federal Government," note 8 above; Robbins, "The Impact of the Delegation Doctrine on Prison Privatization," note 23 above.

[34] See Krent, "Legal Theory: Fragmenting the Unitary Executive: Congressional Delegations of Administrative Authority Outside the Federal Government," note 8 above, at 85–89.

[35] Harold I. Abramson, "A Fifth Branch of Government: The Private Regulators and Their Constitutionality" (1989) 16 *Hastings Const. L.Q.* 165 (pointing out that both the Consumer Product Safety Commission and the Office of Management and Budget have established similar policies of adopting privately generated standards). Ibid. at 173. Influential private standard-setting groups include the American Society for Testing and Materials, the National Fire Protection Association and the American Society of Mechanical Engineers. See Franco Furger, "Accountability and Systems of Self-Governance: The Case of the Maritime Industry" (1987) 19 *L. & Pol.* 447, citing numerous standard-setting organizations.

[36] James W. Singer, "Who Will Set the Standards for Groups That Set Industry Product Standards?" 12 *National Journal* 721 (1980). States and cities also incorporate voluntary standards in building codes.

[37] This was the primary method for establishing OSHA standards in the years after the agency was first created. See Occupational Safety and Health Act of 1970, Pub. L. No. 91–596, 84 Stat. 1590

National Technology Transfer and Advancement Act of 1995 requires federal agencies to use voluntary consensus standards in certain activities as a means of carrying out policy objectives unless the use of those standards would be inconsistent with applicable law or otherwise impractical.[38] The extent of public-private cooperation in standard-setting is substantial; indeed, agencies and standard setting groups could not be more symbiotic. For example, the American National Standards Institute (ANSI) at one point entered a working agreement with OSHA to provide technical support for the development, issuance and application of standards. As part of this effort, OSHA representatives in turn participated on ANSI committees and provide ANSI with information and research reports.[39]

Despite the formal overlay of agency authority and observance of legal procedures, the role that private standard-setting plays in ostensibly public standard-setting might raise doubts about the legitimacy of the resulting regulations. Standard-setting organizations tend to be technically expert, industry-dominated and secretive. They regularly fail to include adequate consumer, small business and labor interests on their technical committees. In addition, the largest firms often exert a disproportionate influence over such organizations, ensuring that standards promote their products.[40] Indisputably, such processes represent opportunities for self-dealing and anti-competitive behavior.[41] In these contexts, concentrated private interests may undermine the larger public interest in product or workplace safety. At the same time, the participation of

(1970) (codified as amended at 29 U.S.C. § 656 (b) 1982). The Act directed the Secretary of Labor to promulgate any national consensus standard unless he determined that it would not result in improved safety or health. See Sidney A. Shapiro and Thomas O. McGarity, "Reorienting OSHA: Regulatory Alternatives and Legislative Reform" (1982) 6 *Yale J. On Reg.* 1, citing 29 U.S.C. § 651.

[38] The National Technology Transfer and Advancement Act of 1995, Pub. L. No. 104–113, 1996 U.S.C.C.A.N. (110 Stat.) 775 (codified in scattered sections of 15 U.S.C.). In addition, the Act requires agencies to participate in the development of voluntary standards when such participation is compatible with an agency's mission, authority, priority and budget resources.

[39] ANSI is a private body that does not write standards itself but operates as a clearinghouse for standards that are set by technical, professional and trade associations. As a result it is enormously influential. For a comprehensive article on private standard-setting, see Robert W. Hamilton, "The Role of Nongovernmental Standards in the Development of Mandatory Federal Standards Affecting Safety or Health" (1978) 56 *Tex. L. Rev.* 1329.

[40] See Singer, "Who Will Set the Standards for Groups That Set Industry Product Standards?" note 36 above, at 722 (claiming that only 14.6% of the membership of the American Society for Testing and Materials is composed of individual members including consumers, ecologists and consultants but that ASTM has financed consumer participation in some of its committees). Both domestic and international standard-setting bodies, such as the International Standards Organization (ISO) are vulnerable to criticism about the absence of balanced representation on committees. Like its domestic counterparts, the ISO's standard-setting process is not broadly participatory and is dominated by technical groups representing industry interests. The American National Standards Institute is the U.S. representative to the ISO.

[41] See Andrew F. Popper, "The Antitrust System: An Impediment to the Development of Negotiation Models" (1983) 32 *Am. U. L. Rev.* 283 (citing *Hydrolevel Corp. v. American Soc'y of Mechanical Engineers, Inc.*, 635 F. 2d 118 (2d Cir. 1981), *aff'd*, 102 S. Ct. 1935 (1982) as an example of anti-competitive behavior by the ASME in the consensual private standard-setting process).

technically expert professional and industry associations might enhance an agency's ability to perform its regulatory function.

The most explicit attempt to include private parties in standard-setting has taken the form of regulatory negotiation, a controversial initiative authorized by federal legislation and used by numerous federal agencies to negotiate consensus-based regulations.[42] Reg-negs, as they are called, are conducted by formally chartered and representative groups which negotiate within parameters established by the agency and according to rules agreed upon by the group. Most reg-negs consume considerable time and resources and require detailed technical knowledge on the part of the participants. Inexperienced parties are almost never invited to participate. Assuming the negotiations culminate in an agreement, the consensus-based rule must still go through the traditional notice and comment process pursuant to the APA s.553. Moreover, even after the participants sign the consensus agreement, the agency is not bound to promulgate the consensus reached.[43]

This form of shared public–private activity attracts criticism because it appears to subject rulemaking to explicit interest group bargaining. The process arouses distrust despite the procedural safeguard of notice and comment because of suspicions that the agency may feel bound to promulgate a consensus with which it does not wholeheartedly agree. Agency officials may, for example, make a policy decision that it is more important to avoid litigating the rule than to insist on a more stringent standard.[44] Alternatively, they may be willing to trade the stringency of the standard for smoother implementation, which they might reasonably anticipate in the wake of a consensus-based agreement. After committing considerable time and resources to a reg-neg, an agency may wish to save face, or it may be so committed to the success of the enterprise that it accommodates private interests more than it otherwise might. Even assuming agency good faith, the fact that only a small number of representatives are selected to participate in reg-negs contributes to the notion that the process amounts to bargaining among powerful players. Excessive reliance on outsiders suggests to some observers that the agency has abdicated its role.

Even the most seemingly independent internal agency processes can rely to a significant extent, however, on outsiders. For example agencies often depend heavily in the standard setting or approval process on advisory panels of scientists and other experts. Such panels may exert an enormous amount informal influence on agency decision making. Their activity is not limited to resolving technical disputes and achieving consensus on the state of science; instead, they often become embroiled in important debates over policy.[45]

[42] Negotiated Rulemaking Act, 5 U.S.C. §§ 561–570 (1994).

[43] *USA Group Loan Services Inc. v. Riley*, 82 F. 3d 708 (7th Cir. 1996).

[44] When parties reach consensus, they typically sign an agreement not to challenge the rule.

[45] Nicholas A. Ashford, "The Role of Advisory Committees in Resolving Regulatory Issues Involving Science and Technology: Experience From OSHA and the EPA" in *Law and Science in Collaboration*, (J.D. Nyhart and Milton M. Carrow, eds., Washington, D.C.: The National Center for Administrative Justice, 1983) at 172. Advisory committees may be permanent, quasi-permanent

Because they are theoretically insulated from private pressure and in the employ of the agency, such panels should raise few accountability problems. Indeed, formally chartered advisory panels to federal agencies are subject to the open meeting and balanced representation requirements of the Federal Advisory Committee Act.[46] And yet, both at the federal and state levels, expert advisory committee members may experience conflicts of interest and be vulnerable to pressure from outsiders.[47] Even financial disclosure requirements cannot prevent interested parties from lobbying panel members or supplying them with flawed or self-interested technical information. Moreover, even when panelists disclose existing conflicts of interest, most of them face conflicts of interest with respect to *future* employment or academic funding opportunities.

In sum, standard-setting relies to a significant extent on public–private cooperation. Both agency incorporation of privately set standards and agency reliance on expert panels arguably warrant greater scrutiny than would standard-setting based *solely* on in-house expertise. The difficulty, of course, is *distinguishing* in-house expertise from dependence on private parties, since private parties are so well integrated into the traditional standard-setting process. These public-private arrangements, whether formal or informal, engender doubts about impartiality, independence, conflicts of interest and self-dealing that remain insufficiently addressed by the mere existence of agency oversight or the application of procedural rules governing private conduct. Imposing even more constraints on *private* actors (producer groups, standard-setting organizations, expert panels) is one way to try to provide more accountability and increase participation, but perhaps at the risk of undermining the special advantages of private contributions.

or ad hoc. They may be broadly representative of diverse interests (OSHA's permanent advisory committee known as the National Advisory Committee on Occupational Safety and Health, for example) or limited to particular kinds of experts (the Environmental Protection Agency's Science Advisory Board, for example). Ibid., at 171–172.

[46] Federal Advisory Committee Act, 5 U.S.C. app. §§ 1–15 (1994).

[47] For these reasons, some commentators are troubled by regulatory reform bills that would require regulations to be vetted by a peer review group of scientists (In the 105th Congress, for example, The Science Integrity Act, H.R. 3234, 1998, was introduced by Representative Pombo "to require peer review of scientific data used in support of federal regulations, and for other purposes" and the Sound Sciences Practices Act, H.R. 2661, 1997, was introduced by Representative McInnis "to establish peer review of standards promulgated under the Occupational Health and Safety Act of 1970."). See Kenneth John Shaffer, "Improving California's Safe Drinking Water and Toxic Enforcement Act Scientific Advisory Panel Through Regulatory Reform" (1989) 77 *Calif. L. Rev.* 1211, 1211 (arguing that the supposedly disinterested expert panel is "a highly politicized and sometimes ineffective body whose decisions seriously threaten both public health and the state's economic welfare."). Notably, state and local advisory committees are not governed by FACA which requires a balanced representation of views. Governors can more easily "stack" committees than can federal officials.

C. "Self-regulation"

1. Voluntary self-regulation

While agencies do frequently adopt and enforce private standards as their own, private standard-setting bodies may also operate independently of, and parallel to, government regulation. Such parallel standard setting might be called *voluntary* self-regulation because, while government might tolerate and even encourage it, regulatory agencies do not necessarily yield their own authority to set and implement standards in the face of self-regulatory systems. Voluntary self-regulation may nonetheless play a powerful role in establishing the de facto standards that govern a particular industry or activity. This depends in part on the government's posture toward self-regulation. For example, an agency may exercise its enforcement discretion favorably when a self-regulating firm technically violates statutory or regulatory standards.[48] The widespread exercise of such discretion to approve ostensibly parallel regulation threatens to turn *private* regulation into de facto *government* regulation, with little public access to the process.

Self-regulatory programs are, in reality, not self-reliant at all. They depend on a network of relationships both within the industry (vertically from the trade association down to members and suppliers, and horizontally across members) and between the industry and public agencies with the authority to influence such measures.[49] A typical self-regulatory initiative in environmental regulation, for example, combines an environmental management system with internal standard-setting or goal articulation, regular audits (sometimes performed by the firm itself, sometimes by independent auditors) and publication of environmental reports (sometimes to regulators only, sometimes to both regulators

[48] Both in the academic literature and in the real world of enforcement, there is a heated dispute over whether firms that engage in self-monitoring and self-auditing should be entitled to special treatment by regulators. Both the EPA and the Department of Justice have adopted policies that allow the mitigation of penalties or exercise of enforcement discretion when firms implement self-monitoring and self-auditing programs. See *Incentives for Self-Policing: Discovery, Disclosure, Correction and Prevention of Violations*, 60 FR 66, 706 (1995); U. S. Department of Justice, *Factors and Decisions on Criminal Prosecutions for Environmental Violations in the Context of Significant Voluntary Compliance or Disclosure Efforts* (July 1, 1991). A separate issue is whether the information revealed in the audits should be privileged and unavailable to either agencies or private actors who wish to use it in enforcement litigation. While numerous states have passed audit privilege legislation, EPA has steadfastly refused to treat audits as privileged and has pressured states to amend their legislation. See Eric W. Orts and Paula C. Murray, "Environmental Disclosure and Evidentiary Privilege" (1997) 1 *Ill. L. Rev.* 1. See also Brooks M. Beard, "The New Environmental Federalism: Can the EPA's Voluntary Audit Policy Survive" (1998) 17 *VA. Envtl. L.J.* 1; Lisa Koven, "The Environmental Self-Audit Evidentiary Privilege" (1998) 45 *UCLA L. Rev.* 1167.

[49] Rees' study of the nuclear industry's self-regulation revealed that the ostensibly private system worked effectively only because the industry's trade association relied on the threat of enforcement by the Nuclear Regulatory Commission. See Rees, *Hostages of Each Other: Transformation of Nuclear Safety Since Three Mile Island, supra,* note 16, at 38–40 (describing how the Institute of National Power Operations ultimately resorted to the NRC's enforcement authority to establish credibility with the industry).

and the public). Most voluntary environmental self-regulation makes no pretense of establishing hard performance standards. Rather, it is designed to inculcate management reforms aimed at "continuous improvement" toward externally imposed norms. Self-regulation does offer a number of benefits to private firms, including market advantage over competitors that fail to meet such standards, lower insurance and loan rates, and improved public relations.

The chemical industry's Responsible Care program is a frequently cited example of such a parallel system. Responsible Care consists of a set of industry codes governing how chemical companies manufacture and distribute their products and interact with their suppliers, distributors and consumers. Because the codes do not impose quantitative performance standards, compliance with them commits firms only to management practices and internal accountability mechanisms (such as auditing and reporting) which are designed to integrate environmental considerations into every aspect of firm decision making, from product design through distribution and sale.[50] The Chemical Manufacturer's Association, the industry's trade association, enforces Responsible Care and has the power to expel non-compliant member firms. The monitoring system is opaque to outside observers because the CMA's auditing process is not publicly accessible. The trade association has never expelled a member firm for non-compliance.[51]

Similar self-regulatory mechanisms arise in the international arena. For example, the Institute for Standards Organization (ISO) in Geneva has published its 14000 series, a set of environmental management standards based largely on the ISO's total quality management standards adopted in its earlier 9000 series.[52] ISO 14000 certification requires firms to assess their environmental effects and establish a management system for achieving continuous improvement.[53] Firms that adopt these standards become "ISO certified," a characterization that produces a number of important economic benefits, including lower insurance or loan rates, access to markets that demand ISO certification, potential market advantage among consumers, and potentially favorable treatment by domestic regulatory agencies. Indeed, domestic companies are likely to feel increasing pressure to adopt ISO standards in order to compete in the global market place.[54]

[50] See Jennifer Nash and John Ehrenfeld, "Codes of Environmental Management Practice: Assessing Their Potential as a Tool for Change" (1997) 22 *Ann. Rev. Energy Environ.* 487 at 499–501.

[51] Ibid.

[52] See Paula C. Murray, "The International Environmental Management Standard, ISO 14000: Tariff Barrier or a Step to an Emerging Global Environmental Policy" (1997) 18 *U. PA. J. Int'l Econ. L.* 577.

[53] Firms seeking certification must, "inventory all of the environmental 'aspects' associated with its activities and products. It identifies those it considers 'significant' and develops a management system that sets targets, allocates resources, provides training of employees and establishes a system for auditing." See Nash and Ehrenfeld, "Codes of Environmental Management Practice: Assessing their Potential as a Tool for Change," note 50 above.

[54] See Murray, "The International Environmental Management Standard, ISO 14000: Tariff Barrier or a Step to an Emerging Global Env'l Policy," note 52 above.

As with Responsible Care, adopting an EMS to satisfy ISO 14000 is not a commitment to achieving specific *performance* standards. Certification guarantees only that a system is in place to meet a firm's goals, but it does not require firms to achieve a particular level of environmental performance. ISO certification is often proposed as an *alternative* to domestic regulatory standards, which impose substantive, technology-based emission and effluent limits as well as process and design standards. Although the EPA has thus far refused to accept ISO certification in lieu of compliance with domestic standards, it has signaled a willingness to exercise enforcement discretion favorably for companies that are ISO certified.[55]

Experiments in voluntary regulation seem to be proliferating, some at the impetus of trade associations representing particular industries (as with Responsible Care),[56] and some initiated by more heterogeneous business networks.[57] As a *substitute* for government regulation, voluntary measures like these are seen as threatening to the legitimacy of environmental regulation. First, as noted earlier, most self-regulatory programs emphasize establishing a management system for integrating environmental decision making into the firm's hierarchy without imposing hard performance standards. Second, while there can be numerous incentives for firms to comply with their own programs, a lack of transparency and public involvement in self-regulation makes it difficult to monitor compliance. While the goal of environmental protection *can* coincide with the private profit motive (where, for example, pollution prevention and waste minimization strategies reduce a firm's consumption of energy and materials, thereby saving costs), sometimes they diverge sharply. At least

[55] EPA's audit policy requires either an EMS such as ISO or a systematic audit in order to obtain some level of relief. See EPA Position Statement on Environmental Management Systems and ISO 14001 at 63 FR 12094 (March 12, 1998).

[56] Examples of trade associations that have embarked on self-regulatory initiatives include the American Forest and Paper Association, the National Association of Chemical Recyclers, the American Meat Institute, the Wisconsin Paper Council, the American Textile Manufacturer's Institute and the American Petroleum Institute (API). I thank Franco Furger for sharing with me his compiled list of trade associations and initiatives. See Nash and Ehrenfeld, "Codes of Environmental Management Practice: Assessing Their Potential as a Tool for Change," note 50 above, at 510, for a description of the API's program, Strategies for Today's Environmental Partnership. See Ibid. at 511, for a description of the American Forest and Paper Association's Sustainable Forestry Initiative. See also, Eric W. Orts, "Reflexive Environmental Law" (1995) 89 *Nw. L. Rev.* 1227.

[57] Business networks are less cohesive than trade associations, and their membership may cut across industries. Examples of initiatives launched by such loosely cohesive networks include the Global Environmental Management Initiative (GEMI), the Coalition for Environmentally Responsible Economies (CERES), Business for Social Responsibility, the International Chamber of Commerce (ICC) and the Social Venture Network. Some of these initiatives are launched by citizen groups in collaboration with companies, others by financial institutions, and still others by business leaders alone. Again, I thank Franco Furger for sharing his compiled list of these initiatives. See Nash and Ehrenfeld, "Codes of Environmental Management Practice: Assessing Their Potential as a Tool for Change," note 50 above, at 503–05, for a description of the ICC Charter on environmental management and GEMI's role in encouraging firms to adopt it. See Ibid., at 512–16, for a description of the CERES principles (designed to encourage disclosure of environmental performance).

some of the time, firms will be tempted to comply with self-regulation only in token ways, co-opt auditors, or be less than forthcoming in their self-audits and environmental reports.

Even when firms expect to receive nothing more than the favorable exercise of enforcement discretion in return for their voluntary compliance with a self-regulatory program, the absence of adequate guarantees about environmental *performance* might be troubling. Especially at a time when courts are reluctant to review the exercise of enforcement discretion, its more frequent use to respond approvingly to voluntary regulation may undermine important public law values.[58] This is especially true when the voluntary program being approved has been adopted with no public input, consists only of environmental management measures with no hard performance requirements, and provides little opportunity for public scrutiny.

2. *Mandatory Self-Regulation*

Mandated self-regulation is an arrangement in which an agency authorizes or requires private actors to self-regulate under the agency's watchful eye. Inevitably, mandated self-regulation is a shared enterprise. The California Cooperative Compliance Program (CCCP) exemplifies such an approach. The CCCP was instituted by OSHA officials in California in an effort to reform the agency's adversarial, enforcement based approach to workplace safety regulation. The program authorized unions and employers to develop and implement workplace safety requirements through collective bargaining.[59] It effectively delegated OSHA's traditional inspection and enforcement role to a joint labor-management safety committee (consisting of two members each from management and labor). The agency pledged not to intervene as long as the program effectively reduced accident rates.[60]

In his nuanced study of the CCCP, Rees explains that both management and labor had strong incentives to cooperate on the safety program. The firms found OSHA's traditional inspection system costly and inconvenient. In addition, they had already begun to pay more attention to safety issues as workers' compensation costs rose. At the same time, unions were motivated to cooperate with employers because they felt increasingly threatened by competition from open shops. More significant than this fear, however, was the presence of an "ideology of cooperation" in construction unionism, which Rees traces to a number of socio-economic factors that have historically bound construction unions symbiotically to their employers.[61]

[58] See *Heckler v. Chaney* 470 U.S. 821 (1985) (holding as unreviewable the exercise of enforcement discretion by the Food and Drug Administration).

[59] See Rees, *Reforming the Workplace: A Study of Self-Regulation in Occupational Safety*, note 15 above.

[60] The program was indeed successful. See Ibid., at 2–3, reporting that accident rates at CCCP projects were "significantly lower" than at other comparable projects in California.

[61] Ibid., at 27–28.

Rees also credits the CCCP's achievements to the power and independence of professional safety engineers within the firms. Their influence had grown in the years prior to the institution of the program, due in part to the initial passage of the OSHA Act. The federal safety mandate had bolstered their organizational status, which positioned them to later play a key role in the CCCP. Worker confidence in the joint inspection committee was another important factor in the program's success. The idea of such a committee built on existing corporate safety programs, so it was not entirely new, but for the first time, the committee enlisted support from employees. Management helped to generate confidence in the new committees by appointing to them knowledgeable and respected employees. The committees in turn proved credible to employees by taking visible and immediate action in response to complaints.

In addition, both labor and management shared important norms, practices and experiences, which facilitated cooperation in an environment with great adversarial potential. They largely agreed, for example, on what constituted a "safety problem." The final ingredient in the CCCP's success was the appointment by OSHA of a designated compliance officer (DCO) for each job site. The DCOs were carefully chosen for their capacity to both monitor and assist in compliance. According to Rees, the DCOs behaved *flexibly*, acting as problem-solving consultants to the process rather than mere enforcement agents.

Although the CCCP example illustrates how mandated self-regulation can work well it implicitly suggests the fragility of such cooperative ventures. They depend on the conjunction of a number of ingredients that are difficult to anticipate at the outset and perhaps impossible to foster. In the absence of a strong union, management might have dominated the standard-setting process, which in turn might have undermined the program's safety goals. Without direct representation on the safety committees, employees would doubtless have been skeptical of the program and would have failed to report accidents. Private firms were not even motivated to control safety-related costs until professional engineers were able to translate worker compensation expenses into a concept of *preventable* accidents. The firms would have been far less likely to actively implement reforms without the help of an internal body of independent professional actors capable of mobilizing support for the "public" goal of safety prevention. Had agency oversight been more remote (had it conformed to the typical OSHA model of occasional inspection), private firms might have been able to escape scrutiny and avoid accountability. Absent a flexible compliance officer skilled in facilitating cooperation, the program might have failed entirely.

D. Traditional Command and Control: Shared Implementation and Enforcement

In an important sense, the traditional command and control regime is already co-regulatory because it relies significantly on private participation in imple-

mentation, even though this is largely hidden from public view. For example, traditional environmental regulation already depends heavily upon industry self-identification for purposes of coverage in a regulatory program,[62] and industry self-reporting for purposes of permit design.[63] Without industry cooperation, most regulatory programs would fail. Private industry has an enormous information advantage over public agencies. Firms provide information to agencies about their own current performance and their potential for improvement given expectations about technology and cost. This information becomes the basis for setting standards, establishing permit conditions and, in the enforcement process, designing consent decrees. Agencies simply lack the resources necessary to do independent research about, properly inspect, and successfully pursue all of the regulated entities that violate regulations.[64]

Self-reporting is, therefore, the heart of the traditional enforcement process, at least in the environmental context. The basis for evaluating industry compliance with permits issued pursuant to the Clean Water Act has always been industry self-reporting about the nature and amount of effluent and the firm's control practices. The same is true for industrial emissions under the Clean Air Act. Frequently, agency staff do not even perform the initial verification of industry reports. Instead, the agency hires contractors to inspect industrial facilities for compliance with regulatory mandates. This practice is ubiquitous in environmental regulation. For example, state and federal agencies also employ contractors to manage clean up of contaminated waste sites under the Superfund remediation process.

In all of these cases, the agency makes the ultimate determination about compliance but essentially removes itself from the front-line of oversight by relying to a significant extent on the regulated entity and third parties. Increased agency dependence on contractors may undermine that ability of agency staff to effectively assess the contractors' judgments, even when the agency retains formal supervisory control over them. By outsourcing enforcement responsibilities to a significant degree, an agency may risk falling below a critical mass of staff or insulating existing staff from direct interaction with regulated firms to such an extent that it severely impairs its supervisory capacity. Moreover, by relying on third party enforcement, an agency risks surrendering control over its enforcement agenda. For example, citizen suits can disrupt an agency's enforcement priorities and ultimately undermine cooperation between private entities and

[62] L. Donald Duke et al., "Widespread Non-Compliance with U.S. Storm Water Regulations for Industry" (1998) in *Environmental Science and Engineering* (in press), (indicating that industry failure to self-identify as covered by the relevant clean Water Act regulations, impedes the effort to regulate storm water runoff).

[63] When firms apply for permits, they provide detailed information about their emissions or effluent, including concentration analysis and flow rate estimation. See Caroline Wehling, "RCRA Permitting" (Winter, 1987) *Nat. Resources & Env't* 27.

[64] See Joel A. Mintz, *Enforcement at the EPA* (Austin: University of Texas Press, 1995) at 114–15. See also Rees, *Reforming the Workplace: A Study of Self-Regulation in Occupational Safety,* note 15 above, at 10.

agencies.[65] Almost all federal environmental statutes provide a private right of action for individuals and groups to sue both the agency (for failure to exercise a non-discretionary duty) and private firms (for statutory violations).

Like regulatory negotiation in the rule making context, EPA's Project XL represents an explicit attempt to directly involve stakeholders in the process of permit design. Project XL grants firms some flexibility in meeting traditional regulatory requirements, providing they equal or exceed applicable environmental standards.[66] A typical facility-based XL project (some are sector-based) might allow a firm to negotiate an agreement in which it makes detailed commitments of "superior" environmental performance in exchange for a more unified, performance-based permit. After receiving approval of its proposed XL project, the firm convenes a stakeholder process to negotiate a Final Project Agreement. Stakeholder groups usually include federal and state agencies, environmental groups and community representatives. The Agreement might authorize the firm to engage in multi-media or cross-pollutant trades (shifting pollutants from one medium to another or trading increases in one pollutant for decreases in another) that would be impermissible under traditional regulations. The Agreement might authorize firms to combine multiple permits with different expiration dates into a single, longer-term permit.[67]

Project XL acknowledges the practical interdependence between regulated industries and regulating agencies in the implementation process. It transforms the traditional arms-length permit negotiation into a more direct and collaborative enterprise. The requirement of a balanced stakeholder group and the participation of relevant agencies is meant to assure that the process provide adequate accountability. Critics of Project XL fault the program for the same reasons they do regulatory negotiation: they believe it allows industry to bargain for beneficial deals that undercut environmental standards established by law and regulation through a process that fails to be sufficiently open and representative. They complain about the uncertainty surrounding the Final Project Agreement, arguing that firms may fall short of their promised "superior" environmental performance and that the cross-media and cross-pollutant trades might seriously undermine the environmental standards firms are supposed to exceed.

[65] Jim Rossi, "Participation Run Amok: The Costs of Mass Participation for Deliberative Agency Decisionmaking," note 9 above.

[66] XL stands for eXcellence and Leadership. The EPA announced that it would consider facilities XL projects, sector-based XL projects and federal facilities XL projects. See Regulatory Reinvention (XL) Projects, 60 FR 27,282, 27,286 and 27282–3 (May 23, 1995).

[67] For a more complete description of Project XL and a close analysis of two projects, see Jody Freeman, "Collaborative Governance in the Administrative State" (1997) 45 *UCLA L. Rev.* 1. For a critical view of XL, see Rena I. Steinzor, "Reinventing Environmental Regulation: The Dangerous Journey from Command to Self-Control" (1998) 22 *Harv. Env'l L. Rev.* 103 at 122–150. For suggestions on improving XL, see Lawrence E. Susskind and Joshua Secunda, "Improving Project XL: Helping Adaptive Management to Work within the EPA" (1998) Working Paper ETP97–02, The Environmental Technology and Public Policy Program, Department of Urban Studies and Planning, Massachusetts Institute of Technology.

On the one hand, the already widespread reliance on private actors in implementation and monitoring suggests that a move to even greater private participation (in the form of programs like Project XL) would be a matter of emphasis rather than a paradigm shift. On the other hand, however, the recognition that reliance on private actors is already pervasive does nothing to alleviate unease about the risks of including private parties more directly in implementation. To the contrary, it suggests that we have been *inadequately* concerned about the accountability problems that attend the traditional command and control system. In other words, Project XL only exacerbates the worst features of an already privately-dominated regulatory process.

III. CONSTRAINING PRIVATE ACTORS

An emerging literature in administrative law suggests that the pressing challenge for the field is to determine when and how to extend legal requirements to private actors performing public functions.[68] In this view, privatization and contracting out disrupt the traditional administrative law project of turning discretion into rules. The trend away from government shifts the administrative law terrain so much that the failure to constrain discretion is decidedly *not* the crucial problem in the field. Instead, the challenge is ensuring that privatization, contracting out, and other measures designed to yield authority to private parties, do not eviscerate the public law norms of accountability, procedural regularity and substantive rationality that administrative law has labored so hard to provide.[69] Viewed in this light, a continued emphasis on constraining agency discretion is like shuffling the deck chairs on the *Titanic*.[70]

Although laudable for its focus on private actors and its bold assertion that agency discretion is no longer the central issue in the field, the emerging privatization literature in administrative law does not go far enough. First, the private role in regulation is even more pervasive and longstanding than the literature

[68] For a collection of essays arguing to this effect, see *The Province of Administrative Law*, note 12 above.

[69] Taggart, "Introduction", in *The Province of Administrative Law*, note 12 above.

[70] That such concerns have arisen first and most forcefully in the United Kingdom, Australia and New Zealand, countries that already impose far fewer legal and procedural constraints on ministerial discretion, and which have witnessed very significant degrees of public sector re-structuring in the last two decades, should not be surprising. See Colin Scott, "The New Public Law" in *Public Sector Reform and The Citizen's Charter* (Chris Willett, ed.; Glasgow: Blackstone Press, 1996); Kerry Jacobs, "The Decentralization Debate and Accounting Controls in the New Zealand Public Sector" (1997) 13 *Financial Accountability and Management* 4, at 331; Graham Scott *et al.*, "New Zealand's Public Sector Management Reform: Implications for the United States (1997) 16 *Journal of Policy Analysis and Management* 3, at. 357. The United States has experienced relatively less privatization largely because a significantly smaller percentage of US industry was publicly owned. Still, there have been some notable signs of reform in the US, which, as with the Commonwealth initiatives, seek to free private activity from excessive government intervention. Examples include airline and utility de-regulation, and increased contracting out of government services small and large.

suggests. Even traditional command and control regulation—a hierarchical arrangement in which the agency dictates and enforces standards—is characterized by significant informal interdependence between government and private actors. Widespread contracting out of public functions and greater reliance on self-regulation might seem to increase the private role, but every aspect of administration is deeply and inevitably interdependent.

The new privatization literature in administrative law is marked by debates over whether judicial review will subside or intensify as the private role in administration increases. Some scholars argue that a proliferation of private activity will weaken the executive and legislative capacity to exert control over public decisions, which will invite greater judicial oversight.[71] Courts may then choose to regulate private actors either by expanding the state action doctrine or by infusing common law doctrines with public law norms, such as good faith obligations in contract. Indeed, there is ample precedent for imposing procedural requirements on private parties performing public functions when they act in derogation of the public interest.[72] Though the relevant common law doctrines have fallen into disuse, they might certainly be revived. As my colleague Michael Asimow recently observed (though with concern rather than approval), the California Supreme Court recently read due process requirements into a private contract between an employer and an employee accused of sexual harassment.[73] This practice may well become more common in the future. Certainly, the question of when private actors ought to be treated as public ones will increasingly occupy administrative law scholars.

In these analyses, scholars will undoubtedly consider the extent to which the benefits of private participation in governance, including expertise, innovation and efficiency, may be frustrated by the imposition of traditional constraints such as compliance with legal procedures and formal accountability to an

[71] See David Mullan, "Administrative Law at the Margins" in *The Province of Administrative Law*, note 12 above. Mullan argues that judicial review might intensify to respond to privatization. The common law has already imported public law notions to govern private behavior, where the activities in question implicate the public interest, raise legitimate expectations of procedural fairness or threaten rights of contract and property. If the private sector continues to exercise more and more public functions, public law values will be applied to those actors and their activities by courts through the common law. Thus de-regulation in the form of government withdrawal or shrinkage, does not mean that the functions performed by different actors will go unregulated. This view might be contrasted with Aman's argument that the new administrative law will largely be determined by the executive and legislature and that courts will play a relatively minor role. See Alfred C. Aman, Jr., "Administrative Law for a New Century" in *The Province of Administrative Law*, note 12 above, at 117.

[72] See Taggart, "The Province of Administrative Law Determined?" in *The Province of Administrative Law*, note 12 above, at 16.

[73] Michael Asimow, "The Private Due Process Train is Leaving the Station" (1998) 23 *Admin. & Reg. L. News* 8, noting *Cotran v. Rollins Hudig Hall International, Inc.*, 69 Cal. Rptr. 2d 900 (1998) (holding that the issue for the jury in a sexual harassment case was whether the employer "reasonably believed" after an appropriate investigation, that the employee was guilty of harassment, not whether he in fact was guilty). Asimow argues that this development threatens to bureaucratize relatively informal relationships.

elected body.[74] Greater participation of private actors in the administrative process may help to produce superior regulatory decisions and facilitate their implementation.[75] Indeed, opportunities for greater participation in governance may have an independent, democracy-enhancing value. There might be circumstances in which we are willing to trade some degree of formal accountability for these other benefits.

The task for administrative law is more complicated, however, than delineating a threshold test to determine when a private actor is performing a sufficiently public function to justify the imposition of public law constraints. Confronted with the reality of shared governance, administrative law scholars must ask new questions. Do private actors have any obligation to be "public-regarding" in setting standards that are then incorporated by reference by an agency, or is the agency's stamp of approval an adequate guarantee of accountability?[76] Should courts review the exercise of enforcement discretion to approve voluntary self-regulation more carefully than when it is exercised for other purposes? May third party beneficiaries of agency-provider contracts sue to enforce their terms? Does the degree to which private parties *already* participate in standard-setting and implementation or the extent to which agencies *already* rely on contractors for enforcement raise accountability concerns that we have been carelessly ignoring?

The necessary inquiry will require highly specific analyses of the dangers (self-dealing, conflicts of interest, secrecy, irrationality, lack of representation, and procedural irregularity, to name some) posed by different regulatory arrangements. Before reflexively imposing traditional constraints on private actors, we need to think carefully about whether they are sufficiently like agencies to justify similar treatment. Moreover, before determining which constraints might be appropriate, we ought to ask whether *other* actors or *different* mechanisms

[74] In his now classic article, Jaffe captured the potential benefits of private participation in governance. "Those performing the operation or constituting a part of the relation to be regulated are likely to have a more urgent sense of the problem and the possibilities of effective solution: experience and experiment lie immediately at hand . . . Participation in management satisfies the craving for self-expression, for power. It is valuable because it may stimulate initiative and quicken the sense of responsibiliity . . . Group self-government democratically organized offers some hope for the development of these qualities in the broad masses of people; it at least suggests that public administrations, superimposed, relatively divorced from the field of operation, and—at least under capitalism—not primarily responsible for results, should not be the exclusive method of regulation." See Louis Jaffe, "Law Making by Private Groups" (1937) 51 *Harv. L. Rev.* 201 at 212. See also David M. Lawrence, "The Private Exercise of Governmental Power" (1986) 61 *Ind. L.J.* 647 at 651–658 canvassing the justifications for delegations to private groups, including pluralism, interest representation, flexibility and expertise.

[75] Empirical evidence suggests that the more involved people are in making rules and consenting to them, the stronger their sense of obligation to abide by them. See Robert Kidder and Craig McEwen, "Taxpaying Behavior in Social Context: A Tentative Typology of Tax Compliance and Non-Compliance" in Jeffrey A. Roth and John T. Scholz, *Taxpayer Compliance, Vol. 2* (Philadelphia: University of Pennsylvania Press, 1989) at 53.

[76] The point of departure for public law is public regarding behavior. See Michael Taggart, "Introduction" in *The Province of Administrative Law,* note 12 above.

might play a role providing accountability and ensuring compliance with other public law norms

Put another way, the impulse to respond to private activity by constraining private actors merely shifts the focus to the private side of the equation rather than re-orienting the administrative law inquiry to the public-private regime as a new entity. Acknowledging the shared nature of governance invites us to explore more fully what we mean when we say that regulation is "unaccountable." What interests lay behind the traditional concern about accountability and how would we weight those concerns in the context of a specific decision making regime that offers some benefits at the cost of other things? The public acceptability or legitimacy of a decision making regime turns in part upon our expectations of how the actors in that regime *ought* to behave when they play certain kinds of roles. For example, when private actors function in an advisory capacity in which they purport to be neutral, we might rightly expect disinterested decision making. Disinterest might matter less, however, in a process like regulatory negotiation, where we might expect parties to pursue their interests (which might, nonetheless facilitate problem-solving that is in the public interest).[77] In this context we might place a premium on participation and adequate representation rather than neutrality. When a private actor plays an enforcement role, either through independent oversight or by exercising a private right of action, we might expect it to behave differently than when it acts in a standard-setting capacity. In the former case, we might worry about private motivations that threaten to derail a rational enforcement agenda. In the latter case, we might want to minimize self-dealing and anti-competitive behavior by ensuring adequate representation on the standard-setting committee of all affected interests.

The imposition of rigorous legal procedures, together with oversight by an elected body that is itself accountable to the electorate, is not the only way to ensure the legitimacy of public-private arrangements. A mixed administrative regime might rely on numerous *informal* accountability mechanisms and *non-governmental* actors to control the dangers posed by public–private arrangements. Indeed, such mechanisms are embedded within, or suggested by, the examples in Part II. Sometimes the legitimacy of a regulatory initiative depends in part on trust and shared norms. Public-private arrangements can be more accountable because of the presence of powerful independent professionals within private organizations or because the agency's threat of regulation provides the necessary motivation for effective and credible self-regulation which itself involves non-government actors. Sometimes the two principal partners in a regulatory enterprise (the agency and the regulated firm) might rely on independent third parties to set standards and oversee enforcement. Even the absence of a direct government role does not mean a seemingly private regime is

[77] While parties clearly pursue their interests in a regulatory negotiation, the deliberative process might help them to alter their initial positions, re-conceive problems and imagine creative solutions. See Freeman, "Collaborative Governance in the Administrative State," note 67 above.

free of regulation or oversight. Informal regulatory regimes can emerge in a context where there is no formal government participation.[78]

Consider the CCCP. The success of failure of a public-private arrangement depends on more than formal delegation of authority and formal oversight. Informal norms and practices about workplace safety, and the trust that they engendered, contributed to the program's success. The professional safety engineers acted independently to check the behavior of management and bolster support for safety initiatives. Strong labor representation on the safety committees also helped to limit the potential for management domination of the cooperative process. Moreover, the CCCP's success cannot be adequately understood without appreciating the "indigenous regulatory factors" that shaped the regulatory environment.[79] These include the socio-economic forces that contributed to an ideology of cooperation in construction unionism, the prior existence of corporate safety programs as a launching pad, the incentives created by a costly workers' compensation system and the galvanizing effect on engineers of the professional safety movement.

The Responsible Care example demonstrates how informal mechanisms can be essential to the effectiveness of self-regulation and at the same time provide some assurance of accountability. The CMA's formal power to enforce Responsible Care may be less important to the program's success than informal disciplinary mechanisms such as peer pressure and institutional norms of compliance. Empirical studies reveal that executives from leading firms pressure their non-compliant counterparts at industry meetings to adopt and adhere to the industrial codes. Publication of the codes has also given leverage to professionals and managers within the industry who wish to take a leadership role in environmental performance.[80]

Responsible Care models how self-regulation can provide an opportunity for experimenting with the most innovative and environmentally protective regulatory strategies. Most self-regulatory systems designed to address environmental problems emphasize technological innovation, life-cycle assessment, benchmarking, continuous improvement and pollution prevention. Indeed, proponents of self-regulation argue that these strategies, which in theory seek to integrate environmental concerns into both every stage of product development (design, distribution and sale) and every business relationship (between firms, suppliers, distributors and customers), have flourished precisely because they were developed by *private* industry. On this view, private actors are sources of innovative regulation – an unsurprising conclusion if true, but one that militates

[78] See Colin Scott, "Analysing Regulatory Space: Implications for Institutional Design and Reform" (unpublished manuscript presented at International Law and Society Conference, Snowmass, June 1998) (describing how British Telecom became the de facto regulator once the British telecommunications industry was de-regulated).

[79] Rees, *Reforming the Workplace*, note 15 above, at 125.

[80] See Nash and Ehrenfeld, "Codes of Environmental Management Practice: Assessing Their Potential as a Tool for Change, " note 50 above. See also Gunningham, and Grabosky, *Smart Regulation*, note 16 above.

toward harnessing self-regulatory efforts rather than prematurely dismissing them as fundamentally unaccountable.

Perhaps most importantly, the example illustrates the perils of generalizing about the threat to accountability posed by self-regulatory initiatives, given the extent to which their features turn on the internal structure of the industry itself and the institutional background against which the self-regulation arises. Responsible Care is widely regarded as the most far-reaching and successful example of a self-regulatory regime. Its success depends, however, on the unique features of the chemical industry, including its relative maturity and stability, its vulnerability to poor publicity and the unusually strong influence of its peak level trade association.[81] As a result, the program may be hard to replicate without a similar convergence of circumstances, structures and relationships.

It would, of course, significantly enhance Responsible Care's credibility were the auditing process independently performed by, or at least subject to, third party verification. In turn, a supervising regulatory agency could play a role overseeing the independent auditors. Third party verification is an increasingly popular tool for ensuring that private firms live up to their voluntary obligations, at least in the environmental arena. For example, the EPA's Environmental Leadership Program relies on third party verification to monitor compliance with existing environmental regulations. The independent third parties are themselves certified by the agency.[82]

Financial institutions—lenders and insurance companies—might also be helpful third party regulators.[83] Again, environmental regulation provides a useful illustration. Until recently, lenders faced potentially massive liability for any toxic waste contamination for which their clients were potentially responsible parties under CERCLA.[84] The threat of exposure motivated lenders to demand stricter environmental compliance from their clients. Along with insur-

[81] See Gunningham and Grabosky, *Smart Regulation,* note 16 above (pointing to the relative maturity of the chemical industry, the power of the trade association and the common interests shared by companies that are all vulnerable to negative publicity).

[82] See George S. Hawkins, "Compliance and Enforcement Changes in Congress and EPA" (1997) 11 *Nat. Res. And Env.* 42 at 4. As part of the ELP, EPA Region 1 has created a program called StarTrack, under which EPA grants certified companies penalty reductions and regulatory flexibility in the form of expedited regulatory decisions and reduced reporting and record-keeping requirements. To obtain certification, companies must be evaluated by independent third parties. Certification requires implementation of an environmental management system, benchmarking to ISO 14000 and the completion of compliance audits.

[83] See Jweeping Er, Howard C. Kunreuther, and Isadore Rosenthal, "Utilizing Third-party Inspections for Major Chemical Accidents" (1997) 18 *Risk Analysis* 145.

[84] Recently, Congress amended section 101(20) of the statute to ease the burden on lenders. The section now makes clear that the "potential" to exert managerial control is an inadequate basis for the imposition of liability to lenders and that lenders who engage in commercially reasonable practices when foreclosing and arranging for the sale of assets are insulated from liability. The amendment followed a disagreement among appellate courts over the test for the imposition of liability on lenders. See *United States v. Fleet Factors Corp.*, 901 F.2d 1550 (11th Cir. 1990) (holding that the test is "potential for control"); *In re Bergsoe Metal Corp.*, 910 F.2d 668 (9th Cir. 1990) (holding that the test is "actual control"); *Kelley v. E.P.A.*, 15 F.3d 1100 (D.C. Cir. 1994) (EPA rule interpreting § 101(20) to require "actual control" invalidated as beyond the agency's discretion).

ance companies, lenders developed programs to help clients adopt environmental management systems and to train employees. The principal objection to reliance on financial institutions instead of public agencies is that they will only discipline private actors to the extent that they are themselves exposed to liability. As the risk subsides (as it did when Congress recently amended the lender liability provisions of the Superfund statute), lenders and insurers will likely retreat from their role as regulators. Moreover, even when they do play an active role in disciplining private firms, the standard of performance demanded will be dictated by the lender or insurer's calculation of risk, rather than a determination of what level of performance would adequately protect the public health. Reliance on private institutions to play such a role raises additional accountability problems to the extent that their processes for determining performance standards are not themselves subject to oversight.

The growth of "informational regulation"[85] could also function as a form of third party monitoring. In environmental regulation, mandatory disclosure requires firms to monitor quantity and quality of emissions and disclose that information to the public and/or public agencies. In some cases, agencies demand that firms provide warnings to the public of toxic exposure or other risks.[86] In the context of contracting out services, greater transparency in the tendering process and better publication of contractual terms between agencies and providers could assist the beneficiaries of the those services in seeking redress for injuries suffered due to breach.[87] To the extent that the informed public encounters high transaction costs and other impediments to collective action (both organizational and cognitive), however, informational regulation may not be an adequate accountability mechanism in the absence of additional, complementary measures. Like the other market measures described here, informational regulation not subject to adequate oversight would likely pose accountability problems itself because of the potential for industry manipulation of the information disclosure process.

Many of these informal, market or other accountability mechanisms might be used simultaneously. The rich institutional context of private prison operation suggests that in addition to public actors (the agency, the legislature and courts),

[85] Paul R. Kleindorfer and Eric Orts, "Informational Regulation of Environmental Risks" (1998) 18 *Risk Analysis* 155. Federal statutes like the Emergency Planning and Community Right to Know Act and state initiatives like California's Proposition 65, require industry to disclose their use of toxic chemicals. Proposition 65 was passed by voter initiative in the November 4, 1986 general election and implemented as the Safe Drinking Water and Toxic Enforcement Act, Health & Safety Code, §§ 25249.5–25249.13. Under its provisions, the Governor must publish a list of carcinogenic and reproductive toxic chemicals. The discharge of these chemicals into sources of drinking water is prohibited.

[86] Richard C. Rich, W. David Conn, and William L. Owens, "'Indirect Regulation of Environmental Hazards Through the Provision of Information to the Public: The Case of SARA, Title III" (1993) 21 *Policy Studies* 16.

[87] On the differences between a highly bureaucratic contractual regime and a more voluntary, less prescriptive one, see Peter Vincent-Jones, "Responsive Law and Governance in Public Services Provision: A Future for the Local Contracting State" (1998) 61 *The Modern L.R.* 362.

private parties and non-traditional mechanisms may play useful oversight roles. For example, lending institutions motivated to protect their investment and insurers wishing to minimize risk may act as third-party regulators over private prison operators. As a condition of the loan or policy, for example, they might require that guards and officials submit to training or that prisons officials develop detailed management plans. As the Supreme Court recently observed, market forces should play at least some role in ensuring that private guards are neither too timid nor too aggressive.[88]

Either the legislature or the supervising agency might facilitate third party participation in oversight by requiring independent monitoring or auditing of prisons by certified professionals. A statute or regulation might stipulate that the prison hire only guards certified by independent training programs. Professionals within the prison (say, medical personnel) might have sufficient institutional power and independence to perform a critical role in maintaining health standards; to insulate them from the wrath of the private employer, such personnel might be hired directly by the state agency. States might also enlist the help of independent prisoner's rights groups by granting them standing to sue for violations of any requirements stipulated in the statute or contract.

In sum, a mix of measures and actors can contribute to the effectiveness and legitimacy of public–private arrangements while minimizing the particular dangers they pose. By the same token, simply because a public entity (the agency) retains ultimate authority over a decision making process may not make that process acceptable or legitimate. When an agency adopts without deliberation a private standard-setting organization's safety standards, for example, we may rightly doubt that the mere fact of agency incorporation ensures the legitimacy of the standards. Formal legal procedures and agency oversight may provide the appearance of adequate accountability, but informal mechanisms and private parties play an important and undervalued role legitimizing public–private arrangements.

IV. TOWARD A CONCEPT OF MIXED ADMINISTRATION

Conceptually, as in reality, agencies are hard to dislodge. The centrality of the agency in scholarly discussion derives from the most basic theoretical and

[88] The Supreme Court recently acknowledged the potential role of market forces in the private prison context, when it held that private prison guards are not entitled to qualified immunity from suit by prisoners charging a violation of the Civil Rights Act, 42 USC § 1983 (1964). See *McKnight v. Rees*, 88 F. 3d 417 (1996), aff'd 117 S. Ct. 2100 (1997). The Court reasoned that, "a firm whose guards are too aggressive will face damages that raise costs, thereby threatening its replacement by another contractor, but a firm whose guards are too timid will face replacement by firms with safer and more effective job records. Such marketplace pressures are present here, where the firm is systematically organized, performs independently, is statutorily obligated to carry insurance, and must renew its first contract after three years. . . To this extent the employees differ from government employees, who act within a system that is responsible through elected officials to the voters and that is often characterized by civil service rules providing employee security but limiting the government department's flexibility to reward or punish individual employees." Ibid. at 2105–108.

doctrinal understandings in public law. Perhaps most fundamental among these understandings is that regulatory power is public power. This stems, in part, from the state-centrism of public law. Federalism principles, which admit little room for non-government actors, undergird administrative law understandings of regulatory power. Federalism scholarship tends to assume a zero-sum model of state and federal power, in which these two levels of government compete for a fixed amount of regulatory power.[89] The key debate in federalism scholarship among constitutional law scholars is over which level of government is legally entitled to exercise regulatory power and under what conditions the Supreme Court should intervene. By contrast, for law and economics scholars, the central question is, which allocation of authority between governments would best enhance efficiency.[90] From either perspective, the actors that count are public actors.

The state of the law and legal commentary on the non-delegation doctrine may also contribute to the perception that agencies are the only legitimate public actors. Rooted in Article I of the Constitution and the Due Process Clause of the Fifth and Fourteenth Amendments, the non-delegation doctrine requires Congress to provide sufficient guidelines when delegating authority to confine the delegates' discretion in implementing their congressional mandates. In the modern era, and at least in terms of federal law, the non-delegation doctrine has fallen into disuse. Virtually any delegation to an agency, no matter how vague, will survive constitutional scrutiny on these grounds.[91] Notwithstanding, its potential to prevent Congress from passing important policy choices on to *agencies*,[92] the Supreme Court has deployed the non-delegation doctrine, in its most

[89] See Stephen Gardbaum, "New Deal Constitutionalism and the Unshackling of the States" (1997) 64 *U. Chi. L. Rev.* 483 (for an argument that the New Deal Court's decisions in a number of areas empowered both federal and state regulatory activity, contradicting the zero-sum view that the federal government gained power at the expense of the states). See also Morton Grodzins, *The American System: A New View of Government in the United States* (Daniel J. Elazar ed., Chicago: Rand McNally, 1966), at 7–8 (invoking the marble cake metaphor to describe American federalism). Inter-governmental relations theorists have criticized the dual federalism model (a separate spheres model that precludes *concurrent* federal and state power) for positing a hierarchical and competitive relationship between the state and federal governments . They fault legal scholars as well for focusing excessively on the constitutional allocation of authority while ignoring how governmental relationships actually work in practice. See Deil Spencer Wright, *Understanding Intergovernmental Relations: Public Policy and Participants' Perspectives in Local, State, and National Governments* (North Scituate, Mass.: Duxbury Press, 1978). While constitutional law scholarship is no longer wed to a dual federalism model, the zero-sum view implicitly persists in debates over the allocation of regulatory authority.

[90] See Richard L. Revesz, "Rehabilitating Interstate Competition: Rethinking the 'Race-to-the-Bottom' Rationale for Federal Environmental Regulation" (1992) 67 *N.Y.U.L.Rev.* 210; Daniel C. Esty, "Revitalizing Environmental Federalism" (1996) 95 *Mich. L. Rev.* 570; Kirsten H. Engel, "State Environmental Standard-Setting: Is There a 'Race' and is it 'to the Bottom'?" (1997) 48 *Hastings L.J.* 271.

[91] But cf., *South Dakota v. U.S. Dept. of Interior*, 69 F. 3d. 878 (8th Cir. 1995), vacated and remanded on other grounds, 117 S. Ct. 286 (1986) (invalidating a delegation to the Secretary of Interior to acquire trust land because of the absence of legislative standards).

[92] Chief Justice Rehnquist has indicated his support for reviving the doctrine to invalidate extremely broad delegations of authority to agencies. See *Industrial Union Dept., AFL-CIO v. American Petroleum Institute*, 448 US 607 (1980).

celebrated cases (albeit six decades ago), to forbid delegation to *private* groups.[93]

Although numerous delegations to private bodies have quietly survived judicial scrutiny since these earlier decisions, private delegations are still likely to be more troubling to courts than even the broadest delegations to public agencies.[94] In future cases, the Court might well invalidate private delegations, especially if the delegated authority touches closely on core public powers.[95] Moreover, state court decisions confirm the idea that private delegations raise judicial concern more than public ones, largely because of unease about anti-competitive behavior and self-dealing among private actors.[96] Indeed, the legal commentary on private delegation tends to underscore the more serious dangers associated with unchecked *private* (as opposed to *public*) discretion.[97] The

[93] Despite widespread scholarly dissatisfaction with the doctrine, the Supreme Court appears unwilling either to apply it meaningfully or abandon it entirely. *See,* David Schoenbrod, "The Delegation Doctrine: Could the Court Give It Substance?" (1985) 83 *Mich. L. Rev.* 1223 (developing and advocating a test to control improper delegation). See also David Schoenbrod, *Power Without Responsibility* (New Haven: Yale University Press, 1993) (arguing that delegation weakens democracy). Cf. Jerry L. Mashaw, "Prodelegation: Why Administrators Should Make Political Decisions" (1985) 1 *J. L. Econ. & Org.* 81 (arguing that broad delegations increase the responsiveness of the administrative system).

[94] See Krent, "Legal Theory: Fragmenting the Unitary Executive: Congressional Delegations of Administrative Authority Outside the Federal Government," note 8 above. Courts appear relatively tolerant of such delegations when the private parties still "function subordinately to the public oversight agency," even when no executive branch oversight exists.

[95] Carter Coal, note 32 above at 311 ("The difference between producing coal and regulating its production is, of course, fundamental. The former is a private activity; the latter is necessarily a government function"). Even Krent, who argues that most private delegations can be accommodated by the Constitution without interfering with the Article II interest in a unitary executive, agrees that some private delegations should *not* be tolerated: ". . . if the delegation outside the federal government is too expansive or touches too closely to areas at the core of executive power, the interest in a unitary executive could still prevail. Consider, for example, Carter Coal. The problem in that case was not merely that private individuals exercised 'public' power, but that the power exercised was so sweeping as to diminish the Executive's control over and accountability for creation and implementation of industry codes." See Krent, "Legal Theory: Fragmenting the Unitary Executive: Congressional Delegations of Administrative Authority Outside the Federal Government," note 8 above, at n.9.

[96] See *State Board of Dry Cleaners v. Thrift-D-Lux Cleaners Inc.*, 40 Cal. 2d 436 (1953); *Coos County v. Elrod,* 125 Or. 409, 267 P. 530 (1928); *Jersey City v. Hague, 18 N.J.* 584, 115 A.2d 8 (1955); *Kenyon Oil co. v. Chief of Fire Dept.,* 15 Mass. App. 727, 448 N.E. 2d 1134 (1983).

[97] Broad delegations to agencies raise different concerns than delegations to private actors. Delegating key policy decisions to agencies without providing sufficient detail to guide agency discretion violates the Article I requirement that Congress make the laws. Private actors are even more unaccountable than agencies, which are at least subject to congressional and executive oversight. Unlike agencies, they operate free of both a formal mandate and an informal institutional tradition of serving the public interest. Moreover, private delegation might serve to enhance Congress' power at the expense of the executive. Article II charges the president with the duty to "take care to execute the laws." Congress aggrandizes its own power by appointing private delegates and insulating them from executive control. See Krent, "Legal Theory: Fragmenting the Unitary Executive: Congressional Delegations of Administrative Authority Outside the Federal Government," note 8 above; Abramson, "A Fifth Branch of Government: The Private Regulators and Their Constitutionality," note 35 above; But Cf. Lawrence, "The Private Exercise of Governmental Power," note 74 above, at 649 (arguing that courts fail to distinguish between the dangers of public vs. private delegations).

natural inference might be that today, courts and commentators are likely to view the *public* exercise of delegated power, no matter how unconstrained by guidelines, as more legitimate than the *private* exercise of public power, which is automatically more suspect because of the private nature of the delegate.

Third, the public–private distinction, so central to constitutional law, continues to be fundamental to administrative law and helps to keep agencies at the center of scholarly inquiry.[98] In constitutional law, state action doctrine determines when private behavior is subject to constitutional constraints. Where private parties perform traditionally public functions that have historically been reserved to the state, courts may consider them to be public actors.[99] In a handful of cases, both American and Commonwealth courts have imposed procedural requirements on private actors by reasoning that they are in effect behaving as public actors.[100] Doctrinal mechanisms like the "source of power" or "public function" tests enable courts to characterize traditionally private actors as public whenever they exercise a sufficiently important and traditionally public regulatory function. This would likely be the starting point for a renewed attempt to impose public law constraints on private actors but it may prove very limited; for numerous reasons, courts appear reluctant to employ the source of power or public function tests beyond narrow bounds.[101]

[98] It is a bedrock principle of constitutional law that constitutional rights operate only against the government. See *Flagg Brothers, Inc. v. Brooks,* 436 U.S. 149 (1978) (holding that the Due Process Clause of the Fourteenth Amendment applies only to the government).

[99] See *Edmonson v. Leesville Concrete Company*, 500 U.S. 614 (1991) (finding state action on the part of a private person exercising peremptory authority to remove two black jurors in a jury trial). In *Leesville*, the Court adopted the approach for determining state action established in *Lugar v. Edmondson Oil Company Inc*., 457 U.S. 922 (1982). The Court asked whether the claimed action resulted from a right or a privilege having its source in state authority and whether the private party charged with the constitutional violation could be described in all fairness as a state actor. To make this determination, the Court in *Leesville* considered the following factors: the extent to which the actor relies on governmental assistance and benefits; whether the actor is performing a traditional government function; and whether the injury caused is aggravated in a unique way by the incidents of governmental authority. Ibid. at 621–22. See also *Lebron v. National RR Passenger Corp.,* 513 U.S. 374 (1995) (Amtrak required to recognize free speech although not a government agency because of government control and performance of government function).

[100] See *R. v. Panel on Take-overs and Mergers, ex parte Datafin plc.,* 1 QB 815 (1987) (a United Kingdom case in which a take-over panel exercising an important regulatory function within a self-regulatory framework was subjected to judicial review). See Murray Hunt, "Constitutionalism and Contractualisation of Government" in *The Province of Administrative Law*, *supra,* note 12 at 28–29 for an analysis of the case.

[101] Despite the hopefulness of some Commonwealth academics, the "public function" test has not been extended beyond the relatively narrow context of the *Datafin* case. Judicial reluctance to expand the American state action doctrine may be due to fears about eroding the fundamental distinction between public and private on which our system of constitutionalism depends. One commentator assessed the Court's restrictive stance on state action as, "congenial to Justices who want to preserve state power against the intrusion of the federal government, and who want to restrict the role of the judiciary in second-guessing the political process." See *The Encyclopedia of the American Constitution* (Leonard W. Levy, editor-in-chief and Kenneth L. Karst, ed.; New York: Macmillan, 1986) Vol. 4 at 1738. For the Supreme Court's important limitation that the government function in question must be *exclusively* reserved to the state, see *Jackson v. Metropolitan Edison Co*., 419 U.S. 345 (1974) (finding no state action even where utility company provides 'essential' public service, because provision of service not traditional public function) and *Flagg Bros*., note 98 above.

Such doctrinal innovations continue, in any event, to rely heavily on the formalistic and conceptually dubious characterization of activity as *essentially* public or private. Indeed, this divide remains resilient in the face of withering attacks from critical legal studies, feminist legal theory, legal postmodernism and outsider legal scholarship. Although state action doctrine is famously inconsistent and difficult to rationalize, there is no question that both the Supreme Court and most legal commentators remain committed to the public–private distinction itself.[102] No matter how blurred the line between public and private and no matter how difficult to design an intellectually defensible test to distinguish them, most scholars agree that there *ought to be* a meaningful difference between the two and that constitutional constraints should apply *only* to the former.

One finds a similar commitment to the public-private distinction in administrative law. State actors are subject to the full panoply of congressional, executive and judicial oversight mechanisms. They must comply with all constitutional requirements, including procedural due process, and unless Congress provides otherwise, the procedural demands of the Administrative Procedure Act (APA).[103] Private actors, by contrast, remain relatively unregulated by procedural norms, except to the extent that their contact with the agency runs afoul of the APA's *ex parte* rules,[104] or contravenes statutes such as the Federal Advisory Committee Act.[105]

Fourth, focusing on private actors as potential partners in governance attracts a visceral skepticism in administrative law. Even if most private delegations survive constitutional scrutiny, there remains significant *cultural* resistance to private bodies playing a formal role in regulation, particularly in the performance of quintessentially public functions, such as standard setting. Any attempt to formally delegate such regulatory powers to private actors would likely encounter significant opposition from those concerned about the potential of private participation to undermine congressional intent or to benefit powerful interest groups at the expense of the larger public interest. Already, rather modest attempts to bring stakeholders more directly into the standard setting and implementation process,[106] have met with a storm of controversy, despite

[102] As one commentator points out, "It is not with high hopes that one turns for guidance to a set of Supreme Court decisions famously dismissed as a 'conceptual disaster area'." See David A. Sklansky, "The Private Police" (1998) 47 *UCLA L. Rev.* quoting Charles Black, "The Supreme Court, 1966 Term—Foreword: "State Action," Equal Protection, and California's Proposition 14", (1967) 81 *Harv. L. Rev.* 69, 95.

[103] Administrative Procedure Act, 5 U.S.C. § 553 (1994).

[104] Section 557 of the APA forbids *ex parte* contact by agency officials with any interested party in a formal rulemaking or adjudication.

[105] 5 U.S.C app. §§ 1–15 (1994). The Federal Advisory Committee Act structures governmental consultation with private groups and requires, with some exceptions, that federal agencies formally charter groups with which they meet to obtain advice. The statute requires that the General Services Administration and Office of Management and Budget approve of such charters. A balance of views must be represented in the group.

[106] These efforts have been adopted by the executive, (see Clinton's regulatory reinvention initiatives, including Project XL, the Common Sense Initiative and the Environmental Leadership

numerous procedural checks on such processes, including the reservation of the ultimate decision making authority in the agency.[107]

It would be naïve to quarrel with the concern about agency capture. Chastened by practical experience with powerful regulated industries and influenced by public choice theory, administrative law has grown sensitive to the excesses of pluralism.[108] Public choice theory presumes that private interests (be they firms or "public interest" organizations, labor unions, trade associations or consumer groups) are rent-seekers bent on pursing their interests at the expense of the larger public interest.[109] Indeed, the strongest version of the public choice claim resists altogether the notion of a public interest. Rather, regulation is the product of deal-making between private actors able to provide rewards to legislators and bureaucrats motivated by the desire for job security or other forms of personal gain. Of course, the public choice account of agency action competes with alternative explanations in which, for example, the agency acts as a neutral expert or reaches decision only after engaging in "public-regarding" deliberation over the public interest.[110] Nonetheless, much legislation and many regulations can be explained in public choice terms. Although it strikes some commentators as a cynical theory with potentially corrosive effects, public choice is grounded in powerful economic models and offers a compelling thesis.

Whether or not one subscribes to the public choice view of legislators, administrators and interest groups, the assumptions about private behavior that characterize public choice theory exemplify the relatively truncated view of private participation that dominates administrative law. Even those who resist public choice explanations as too extreme tend to think that private parties play a narrow and mostly rent-seeking role in governance. Given the weaknesses of the extreme public choice explanation, however, the extent to which administrative

Program), endorsed by Congress (see, for example, the Negotiated Rulemaking Act, 5 U.S.C. §§ 561–70, 581, 582–90), supported by the now defunct Administrative Conference of the United States, and cheered by a handful of scholars. See Freeman, "Collaborative Governance in the Administrative State," note 67 above; Philip J. Harter, "Negotiating Regulations: A Cure for Malaise" (1982) 71 *Geo. L. J.* 1.

[107] See William Funk, "Bargaining Toward the New Millenium: Regulatory Negotiation and the Subversion of the Public Interest" (1997) 46 *Duke L. J.* 1351; Susan Rose-Ackerman, "American Administrative Law Under Siege: Is Germany a Model?" (1994) 107 *Harv. L. Rev.* 1279. See also USA Group Loan Services, note 43 above.

[108] See Mashaw, *Greed, Chaos and Governance,* note 3 above at 22 (citing as examples of capture the Interstate Commerce Commission's regulation of the railroads, the Federal Communication Commission's regulation of the broadcast industry and the Federal Power Commission's regulation of the natural gas industry).

[109] From a public choice perspective, legislation can be best understood as a deal struck between private interests and self-interested lawmakers. Agencies are merely another forum in which private interests bargain over the implementation of the "deal" and courts cement those deals in interpreting them. See Gary S. Becker, "A Theory of Competition Among Pressure Groups for Political Influence" (1983) 98 *Q. J. Econ.* 371; Jonathan R. Macey, "Promoting Public-Regarding Legislation through Statutory Interpretation: An Interest Group Model" (1986) 86 *Colum. L. Rev.* 223.

[110] See Mark Seidenfeld, "A Civic Republican Justification for the Bureaucratic State" (1992) *105 Harv. L. Rev.* 1511.

law conceives of private actors *exclusively* in this light is surprising.[111] While private actors undoubtedly pursue their interests, this hardly captures the nuances of their pervasive role in governance. A more complete description of private participation in the administrative process might temper or add new dimensions to the public choice view of private groups. Instead of orienting administrative law *solely* toward erecting barriers to private participation in order to insulate legislators and administrators from influence, we might explore how to harness private capacity in the service of public goals. Without challenging the exclusive focus on rent-seeking in the extreme public choice account, this proves difficult to do.

All of these conceptual and doctrinal forces—a hierarchical and government-centered federalism, the non-delegation doctrine, the public–private distinction, concern about the excesses of pluralism—contribute to the persistent notion that agencies are the only legitimate regulatory actors. Doubtless there are other important and good reasons why the field remains so focused on agency discretion. Recognizing the deep roots of the agency emphasis helps to explain why shifting the inquiry to shared governance and its implications might prove challenging indeed.

V. CONCLUSION: NEW DIRECTIONS IN ADMINISTRATIVE LAW

The concept of mixed administration would be incomplete without a normative theory of the state's role in such a regime. Equally important is developing a companion normative theory of judicial review. Although an adequate treatment of these enormously important subjects is clearly beyond the scope of this essay, I will risk a few preliminary remarks. Shared governance demands a flexible and facilitative notion of government. By this I mean that the state must have the capacity to play a variety of roles in a mixed regime: broker, networker, supervisor, enforcer and partner, to name a few. The state's primary role in a mixed regime may be to *facilitate the intervention of whichever combination of actors proves best capable of maximizing the benefits and minimizing the dangers posed by any particular public-private arrangement*. The suggestions for what a legislature might do in the private prison context to stimulate third party oversight and independent certification are suggestive. In addition, government might creates rights of private action to sue providers for breach of contract, fund public interest groups to participate in regulatory negotiations, promote information disclosure, demand that private standard setting groups diversify their committee membership, and supply government "consultants" to self-regulatory processes.

[111] See Mashaw, *Greed, Chaos and Governance*, *supra,* note 3 (arguing that numerous examples of legislation and administrative action defy explanation on public choice grounds alone). See, generally "Theory of Public Choice Symposium" (1988) 74 *VA. L. Rev.* 167.

Judicial review in a shared governance regime should encourage public–private cooperation while providing a check on the dangers such arrangements pose. Of course, this broad prescription is more a directional impulse than a helpful guiding principle. As Aman rightly notes, "For these approaches to evolve into a new administrative law . . . it is necessary for courts to provide the doctrinal flexibility to incorporate new mixes of the public and the private without, necessarily, opting for one extreme or the other."[112] At a minimum, judicial hostility to processes like regulatory negotiation and continued adherence to a rigid public–private divide, do not bode well for such a transition.[113] Whatever test might be used to determine when public–private arrangements are subject to judicial review (and this will be a matter of considerable debate) the *content* of the review must include an inquiry into non-traditional accountability mechanisms. This suggests an inquiry closer to the pre-*Chevron*, multi-factor test for reviewing agency action than to *Chevron* deference.[114] The critical difference, of course, is that the decision making institution subject to review is no longer necessarily the agency. In addition, as part of the inquiry, courts would be considering factors that are not normally part of the traditional analysis when reviewing agency action, including the availability non-traditional accountability mechanisms.[115]

Notwithstanding the importance of developing these theories, they are secondary tasks. Administrative law must first significantly re-orient itself to study the complicated public-private arrangements that characterize contemporary regulation. Other scholars have made sympathetic, though slightly different, appeals in the past. For example, in an essay reviewing Edley's *Administrative Law*, Rose-Ackerman suggests that the most significant problem with contemporary regulation is the frequent failure to achieve public policy goals, rather than the nagging problem of scope of review.[116] Rose-Ackerman is one of the few legal scholars to point in the direction of studying non-agency actors in an effort to focus administrative law on the "substantive realities of administration."[117] Her critique of Edley merits quoting at length:

> His focus, in typical separation of powers style, is entirely on government. Yet the interaction between private organizations and the government is of central importance in understanding the development of the modern administrative state. Agencies typically contract out for many of their scientific tasks and use private organizations

[112] Aman, "Administrative Law for a New Century," note 71 above, at 117.

[113] USA Group Loan Services, note 43 above.

[114] *Chevron USA, Inc. v. Natural Resources Defense Council, Inc.*, 467 U.S. 837 (1984).

[115] Of course, administrative law scholars will likely balk at this reversion to contextual analysis. Courts might resist it as well, especially the Supreme Court, which may have believed that Chevron was a useful tool for controlling its docket. See Peter L. Strauss, "100 Cases a Year" (1987) 87 *Colum. L. Rev.* 1093.

[116] Christopher F. Edley Jr., *Administrative Law: Rethinking Judicial Control of Bureaucracy* (New Haven: Yale University Press, 1990).

[117] Susan Rose-Ackerman, "Triangulating the Administrative State, A Review of Administrative Law: Rethinking Judicial Control of Bureaucracy by Christopher Edley, Jr." (1990) 78 *Cal. L. Rev.* 1415, 1419.

> to administer programs and provide services. Even the judiciary makes use of special masters who are lawyers . . . A broader attempt to break down the traditional categories of administrative law would need to recognize the role of the private sector in performing the roles of factfinder, policymaker, and administrator.[118]

In another recent attempt to re-direct administrative law scholars, Rubin proposes that the field turn to a "micro-analysis of institutions . . . aimed at the practical problems of governance and the institutions that might solve these problems."[119] Like Rubin, I seek to revive the project of investigating the suitability of different institutions for solving social problems,[120] but with one significant modification. There is no reason to limit micro-institutional analysis to the three institutions at the heart of legal process theory: courts, the legislature and the executive. Seemingly as an afterthought, Rubin suggests that private firms be added to this list. Not only should private firms be a subject of institutional analysis, however, so too should non-profits, trade associations, financial institutions and the host of private actors that already perform, or could perform, significant roles in a mixed administration regime.[121]

Finally, in the absence of careful consideration, administrative law scholars should resist the impulse to constrain private actors as if they were agencies. I say this not only because private activity yields important benefits such as expertise, innovation and efficiency that might be frustrated by the imposition of traditional constraints.[122] Nor do I advance this view merely because private actors are often *already* adequately constrained by their own procedures or by other informal and formal mechanisms. Instead, the impulse to constrain private actors with traditional means may be misguided to the extent that it continues to assume that private and public actors are distinct and that we can easily disaggregate their administrative roles in order to discipline one or the other.

Indeed, virtually all of the legal commentary on private participation in regulation or service provision assumes a clear distinction between public and

[118] Susan Rose-Ackerman, "Triangulating the Administrative State, A Review of Administrative Law: Rethinking Judicial Control of Bureaucracy by Christopher Edley, Jr." (1990) 78 *Cal. L. Rev.*, at 1418–1419.

[119] Edward L. Rubin, "The New Legal Process, The Synthesis of Discourse, and the Micronalysis of Institutions" (1996) 109 *Harv. L. Rev.* 1393. In his article, Rubin makes the case for the viability of this project by arguing that it is compatible with the two dominant theoretical trends in legal academia, outsider scholarship and law and economics, and that it promises to pursue their separate but overlapping ambitions.

[120] Ibid.

[121] Rubin's use of the term micro-institutionalism describes an *inquiry* rather than a *method*. A number of different methodologies might be brought to bear on the micro-institutional project, including those drawn from economics (particularly game theory), political science (particularly implementation studies) and sociology (particularly insitutionalism). See John T. Scholz, "Cooperative Regulatory Enforcement and the Politics of Administrative Effectiveness" (1991) 85 *Am. Pol. Sci. Rev.* 115 (applying interest group theory to implementation analysis). By institutionalism, I refer to the school of thought in sociology that is concerned with informal power—routines, beliefs, reward systems and discourses—that, together with formal legal authority, structure behavior, create norms and engender meanings. Law helps to structure the relationships and incentives that drive organizations, but it is only one among other influences.

[122] See Lawrence, "The Private Exercise of Governmental Power," note 74 above.

private actors. This image of watertight compartments belies reality. In fact, as I have sought to show, administration is an enterprise characterized by interdependence among a host of different actors (agencies, private firms, lenders, insurers, customers, non-profits, third party enforcers, and professional associations, for example). Government and non-government actors operate in a context of institutional richness, in relationship to each other, and against a background of legal rules, informal practices and shared understandings. These *public–private arrangements* defy easy division into purely public and purely private roles.

Most significantly, then, an inquiry into shared governance may ultimately disrupt our notion of the categories "public" and "private" which are so fundamental to administrative law. This goes one very important step beyond merely recognizing that private actors play important public roles or arguing over whether, when they do, they ought to be subject to the same constraints as agencies. It goes further still than the notion that on some occasions and in some contexts private actors might independently provide adequate accountability and comply with public law norms. The degree of interdependence between agencies and non-agency actors may be so complete that it blurs the line between the two categories. If this is true, it no longer makes sense to talk in terms of one or the other. We might substitute instead the notion of *regulatory regimes* in which agencies are in dynamic relationships with private actors.[123]

Understood in this light, the traditional pre-occupation with controlling agency discretion is suspect. If we are concerned about accountability (and administrative law scholars are), then the question is not how to make *agencie*s accountable, but how to make *regulatory regimes*—by which I mean decision making processes and institutions *of which agencies are a part*—accountable. If we are concerned about other public law values such as impartiality, rationality, participation and even efficacy, than we must assess these regimes in terms of their ability to deliver those outputs. Either way, the unit of analysis is *different*. That we have yet to describe these public-private arrangements in new and more analytically useful terms represents both a glaring failure of theory and an exciting opportunity.

[123] For a sympathetic non-hierarchical view of regulation and regulatory actors, See Scott, "Analysing Regulatory Space: Implications for Institutional Design and Reform," note 78 above. Scott's conception of regulatory space occupied by a variety of public and private actors, each with different resources, contrasts with a state-centered conception of regulation in which the public agency has formal top-down control over standard-setting, implementation and enforcement.

Index