

Rules and Principles: A Theory of Legal Certainty

JOHN BRAITHWAITE*

The theory advanced is that precise rules more consistently regulate simple phenomena than principles. However, as the regulated phenomena become more complex, principles deliver more consistency than rules. A central reason is that the iterative pursuit of precision in single rules increases the imprecision of a complex system of rules. By increasing the reliance we can place on a part of the law we reduce the reliability of the law as a whole. Then it is argued that consistency in complex domains can be even better realised by an appropriate mix of rules and principles than by principles alone. A key choice here is between binding rules interpreted by non-binding principles and non-binding rules backed by binding principles. The more complex the domain, the more likely it is the latter that will deliver greater consistency. Robert Baldwin argues that the reason “Why Rules Don’t Work” is that they are typically evaluated without reference to the context of their implementation. Hence we cannot understand when law is and is not consistently implemented by the police without confronting the fact that police culture is not a rulebook, but a storybook. In complex domains, when police, regulatory inspectors and judges enforce rules consistently, they do so as a result of shared sensibilities. Regulatory conversations that foreground obligatory principles buttressed by non-binding background rules are hypothesised to be the stuff of legal certainty on such complex terrain.

This essay conceives rules as specific prescriptions, principles as unspecific or vague prescriptions. In much American legal writing the “rules versus standards” debate is grounded in exactly the same distinction.¹ Is the rule of

* Law Program, Research School of Social Sciences, Australian National University. This paper was completed while I was the Meyer Visiting Research Professor at New York University Law School. My thanks to Reuven Avi-Yonah, Joseph Bankman, Julia Black, Stephen Bottomley, Valerie Braithwaite, Peter Drahos, Rick Krever, Fred Schauer and participants at the Australian Society of Legal Philosophy Conference in Canberra, 2001 for comments on this paper and to Monica Garcia for research assistance.

¹ See this usage of rules and standards in Frederick Schauer, ‘The Convergence of Rules and Standards’, *Regulation Policy Program Working Paper RPP-2001-07* (Center for Business and Government, John F. Kennedy School of Government, Harvard University, 2001) (standards are simply “vague, less specific rules” at 2). I prefer to distinguish standards in

law necessarily a rule of rules or can it be a rule of principles? For some influential lawyers law means quite simply decisions according to rules. One is US Supreme Court Justice Antonin Scalia: "A government of laws means a government of rules. Today's decision on the basic issue of fragmentation of executive power is ungoverned by rule, and hence ungoverned by law."² Philip Selznick conceives the crux of the rule of law in a more complex way to be the restraint of state power by "rational principles of civic order."³ Principles are important on this view that "...the proper aim of the legal order, and the special contribution of legal scholarship, is to minimize the arbitrary element in legal norms and decisions."⁴ Since Aristotle it has been understood that precision in this pursuit can be self-defeating: "our discussion will be adequate if it has as much clearness as the subject matter admits of, for precision is not to be sought for alike in all discussions..."⁵

Aristotle's perspective has more force today. My own interest in the choice between rules and principles in the rule of law has been piqued by the empirical finding in my study with Peter Drahos that global business regulation is a rule of principles rather than a rule of rules.⁶ We exemplified the reason for this by suggesting the absurdity of a state negotiating the harmonization or mutual recognition of telecommunications law by thumping its national telecommunications act on the table as its negotiation position. Because of the impossible complexity if all states did this, what states actually do in global fora is negotiate from principles.⁷ In *Global Business Regulation*⁸ we argue that not only does cannot imply ought not for a rule of rules, there are reasons to do with salvaging deliberative democracy in global decisions why a rule of principles is normatively desirable.

the way business and voluntary standardization bodies do so; standards are norms written in a way to measure performance (see John Braithwaite and Peter Drahos, *Global Business Regulation* (Cambridge Univ. Press, Cambridge, 2000) 19-20. But this distinction is not relevant to the analysis in this essay, which is only about the distinction between specific prescriptions (normally called rules) and unspecific prescriptions (variously referred to in the literature as principles, standards or general rules).

² *Morrison v Olson*, 108 S Ct 2597 (1988).

³ Philip Selznick, *Law, Society and Industrial Justice* (Russell Sage Foundation, 1969) at 11.

⁴ *Id* at 13.

⁵ Aristotle, *Nicomachean Ethics*, Bk I, ch 3 at 1094b, 13-14 (W Ross trans. 1940).

⁶ John Braithwaite and Peter Drahos, above n 1.

⁷ John Braithwaite and Peter Drahos, above n 1 at chs 21 and 24.

⁸ *Id*.

Putting aside complications of globalisation that strengthen the arm of the Aristotelian approach, I will still find virtue in it within the terms of the mid-twentieth century tradition of jurisprudence. This tradition proceeds from the assumption that law is written and interpreted by national legislatures and judiciaries. At one extreme of rule of law scholarship is Hayek, who wants citizens to be free from state interference that is not specified in a rule.⁹ There may be a normative continuum here¹⁰ from libertarians such as Hayek who subscribe to a rule of rules to liberals such as Dworkin¹¹ to civic republicans such as Sunstein¹² and critical scholars such as Kennedy.¹³ For republicans conflict over “incompletely specified agreements” has democracy-building uses: “People who are able to agree on political abstractions – freedom of speech, freedom from unreasonable searches and seizures – can also agree that they are embarking on shared projects. These forms of agreement help constitute a democratic culture.”¹⁴

This essay does not seek to resolve this contest between libertarian, liberal, republican and critical legal studies visions of the rule of law. It simply seeks an empirical understanding on a more specific question that should inform this normative choice. That question is: What are the conditions where rules will deliver us more legal certainty and what are the conditions where principles will do so? While it is a delimited facet of the rule of law, some would see it as a central one, for example Martin Krygier,

⁹ Friedrich A von Hayek, *The Constitution of Liberty*, (University of Chicago Press, Chicago, 1960) 220-35.

¹⁰ Yet Fred Schauer is undoubtedly correct in commenting on this paper that it is wrong to infer from this kind of continuum that rules are inherently a tool of the right and principles of the left. This continuum represents a contingent, though hardly universal, feature of contemporary jurisprudence. Indeed my own conclusion will be that wherever one lies on any political continuum, both rules and principles will be vital tools of regulation.

¹¹ Ronald Dworkin, *Law's Empire* (Harvard University Press, Cambridge, Mass, 1986).

¹² Cass Sunstein, *Legal Reasoning and Political Conflict* (Oxford University Press, Oxford, 1996).

¹³ Duncan Kennedy, ‘Form and Substance in Private Law Adjudication’ (1976) 89 *Harvard Law Review* 1685.

¹⁴ *Id* at 36. Actually, there is not a large difference from the Dworkinian version of liberalism on this score. Dworkin’s “law as integrity” asserts that “people are members of a genuine political community only when they accept that their fates are linked in the following strong way: they accept that they are governed by common principles, not just by rules hammered out in political compromise. Politics has a different character for such people. It is a theatre of debate about which principles the community should adopt as a system, which view it should take of justice, fairness, and due process...” Ronald Dworkin, above n 11 at 211.

for whom Lockean rule of law values “above all seek to ensure that power cannot catch us unawares.”¹⁵

Certainty, consistency and reliability will be used interchangeably here. As in scientific measurement, reliability – getting the same result each time – does not mean that the outcome is correct. A miscalibrated 12-inch ruler that is in fact 13 inches long will consistently measure the length of objects, but it will be consistently wrong. This is an essay on the reliance citizens can put in the consistency of rules, not a contribution to the conditions in which rules provide morally correct solutions to problems.

In legal scholarship it is mostly assumed that tightly specified rules increase legal certainty. For example, in Colin Diver’s leading works on regulatory precision, his vagueness critique is that imprecision results in “indeterminacy”: “Vagueness is a common affliction of regulatory standards, especially those that rely on such open-ended terms as ‘in the public interest’, ‘feasible’, or ‘reasonable’”.¹⁶

What is the Distinction between Rules and Principles?

Ronald Dworkin sees rules as “applicable in an all-or-nothing fashion” when they are crafted to exhaustively include all of their exceptions:

If the facts a rule stipulates are given, then either the rule is valid, in which case the answer it supplies must be accepted, or it is not, in which case it contributes nothing to the decision.¹⁷

Dworkin, in contrast, sees legal principles as not setting out legal consequences that follow automatically when the conditions provided are met. A principle states a reason that argues in one direction, but it does not prescribe a particular decision. Because principles have less specificity in this way, unlike rules principles can conflict. Decision makers assign weights to principles to resolve such conflicts: “it is an integral part of the concept of a principle that it has this dimension, that it makes sense to ask how important or how weighty it is”.¹⁸ Joseph Raz¹⁹ and

¹⁵ Martin Krygier, ‘Ethical Positivism and the Liberalism of Fear’ in T Campbell and J Goldworthy (eds), *Judicial Power, Democracy, and Legal Positivism* (Ashgate, Aldershot, 2000) 64.

¹⁶ Colin S Diver, ‘Regulatory Precision’ in K Hawkins and J Thomas (eds), *Making Regulatory Policy* (Uni of Pittsburgh Press, Pittsburgh, 1989) 200. See also Colin S Diver, ‘The Optimal Precision of Administrative Rules’ (1984) 93 *Yale Law Journal* 65.

¹⁷ Ronald M. Dworkin, ‘The Model of Rules’ (1967) 35 *University of Chicago Law Review* 25.

¹⁸ *Id* at 27.

Frederick Schauer²⁰ have attacked Dworkin's basis for the distinction. They argue that the logical difference between rules and principles has nothing to do with the possibility of conflict or the ways such conflicts are resolved. For Raz, "Rules prescribe relatively specific acts; principles prescribe highly unspecific actions".²¹ A legal principle of environmental regulation like "continuous improvement" can imply an infinitely creative range of action possibilities; a rule preventing the dumping of chemical X relates only to that action. This much is common ground between Raz, Schauer and Dworkin and is also a conception "which accounts for the non-legal use of these terms."²² For the purposes of this essay, we focus on this common ground and ignore the disputed basis of the distinction in the writing of Raz, Schauer and Dworkin. An important implication of this core difference for the practice of jurisprudence is that for both Raz and Dworkin:

A court can establish a new rule in a single judgment that becomes a precedent. Principles are not made into law by a single judgement; they evolve rather like a custom and are binding only if they have considerable authoritative support in a line of judgments. Like customary law, judicially adopted principles need not be formulated very precisely in the judgments that count as authority for their existence. All that has to be shown is that they underlie a series of courts' decisions, that they were in fact a reason operating in a series of cases.²³

The attacks upon positivism, for example from legal realism²⁴ and critical legal studies,²⁵ for indeterminacy/arbitrariness can be responded to by adopting a very "loose" conception of rules that collapses the distinction between principles and rules in the fashion proposed by Robert Goodin.²⁶ For Goodin, rules and principles define opposite ends of a continuum:

¹⁹ Joseph Raz, 'Legal Principles and the Limits of Law' (1972) 81 *Yale Law Journal* 823.

²⁰ Frederick Schauer, 'Prescriptions in Three Dimensions' (1997) 82 *Iowa Law Review* 914.

²¹ Raz, above n 19 at 838. The specificity of rules is fairly common ground among leading positivists. For example, Tom Campbell emphasises that rules "must not be general in the sense of being vague or unspecific". "[S]pecificity, clarity and mutual consistency" are seen as things that automatically go together. Tom Campbell, *The Legal Theory of Ethical Positivism* (Dartmouth, Brookfield, VT, 1996) 6.

²² Raz, above n 19 at 834.

²³ Id 848.

²⁴ For example, Karl N Llewellyn, *Jurisprudence: Realism in Theory and Practice* (University of Chicago Press, Chicago, 1962).

²⁵ For example, Mark Tushnet, 'Following the Rules Laid Down: A Critique of Interpretivism and Neutral Principles' (1983) 96 *Harvard Law Review* 781.

²⁶ Robert E Goodin, *Political Theory and Public Policy* (University of Chicago Press, Chicago, 1982) ch 4.

‘Principle’ is to ‘rule’ as ‘plan’ is to ‘blueprint’, the latter being merely a more detailed form of the former in each case.”²⁷ Confronted with such a move, I would want to persist with the claims in this article, but reframe them to apply to rules conceived along such a continuum. This means restating the key claim in the Abstract as a hypothesis about a set of continuous variables:

As the complexity, flux and the size of regulated economic interests increase, certainty progressively moves from being positively associated with the specificity of the acts mandated by rules to being negatively associated with rule specificity.

Regulating Simple, Stable Phenomena without High Stakes

Some positivists defend themselves against the indeterminacy attacks of legal realists and critical legal scholars by asserting that mostly rule-based judgments are determinate. And indeed it is right to point out that the focus of critical legal theorists on hard cases adjudicated in appellate courts exaggerates the indeterminacy of rules. Outside the courtroom, we get out of bed each morning performing a dozen actions in the preparation of our breakfast that clearly do not breach any legal rule. As we drive to work and exceed the speed limit we clearly do breach a legal rule and so on through the day it is abundantly clear for nearly all of our actions whether they do or do not breach a legal rule. Most of the things we do in life are simple, stable patterns of action where simple, stable rules can provide us with admirable reliance as to what our legal obligations are.²⁸ Hence, the first proposition of the theory of legal certainty advanced here is:

1. When the type of action to be regulated is simple, stable (not changing unpredictably across time) and does not involve huge economic interests, rules regulate with greater certainty than principles.

This is important to note, but unfortunately most matters that become disputes in the courts are rather unlike the actions we undertake when we prepare our breakfast. The first feature that distinguishes the majority of disputes in appellate courts is that for most of the last century they have not involved individual action at all, but corporate action.²⁹ This means that

²⁷ Id at 63.

²⁸ See Ken Kress, ‘Legal Indeterminacy’ (1989) 77 *California Law Review* 283.

²⁹ See the discussion on the problem that Western law is conceived in a tradition that applies to disputes between individuals when the empirical evidence of recent decades is that most disputes in the appellate courts involve corporate litigants on at least one side in John Braithwaite,

they have a degree of complexity – organizational complexity, which, we will see means it is often difficult to know who, if anyone, perpetrated what action. Moreover, corporate life in a post-industrial society is buffeted by technological change and flux in the flow of global markets that violates the stability condition of proposition 1 above. Finally, when corporations fight in the courts, a lot of money is usually at stake. If there are not serious dollars at issue, rational corporations are usually willing to settle or let customers, workers or other disputants win for the sake of building good business relationships and avoiding litigation costs that are higher than the small stakes in the dispute. Hence, while proposition 1 is significant, so are the remaining hypotheses of the theory I wish to advance.

In the next section we move on to consider the second of these: that principles are more likely to enable legal certainty than rules when complex actions in changing environments and considerable economic interests are at stake. Our theory development method will be inductive, relying particularly heavily on three domains of empirical regulatory research: the work of the Centre for Tax System Integrity of the Australian National University plus the work that one of its members, Doreen McBarnet has undertaken with Christopher Whelan on creative compliance in the UK³⁰; the work of the Nursing Home Regulation in Action research group at the ANU and Julia Black's work on financial services regulation in the UK in *Rules and Regulators*.³¹

2. *With complex actions in changing environments where large economic interests are at stake principles are more likely to enable legal certainty than rules.*³²

This hypothesis directly confronts the intuitions of most lawyers. Raz articulates this intuition clearly, without providing any empirical evidence for it: "Principles, because they prescribe highly unspecific acts, tend to be

Restorative Justice and Responsive Regulation (Oxford University Press, Oxford, 2002) ch 8.

³⁰ Doreen McBarnet and Christopher Whelan, 'The Elusive Spirit of the Law: Formalism and the Struggle for Legal Control' (1991) 54(6) *Modern Law Review* 848.

³¹ Julia Black, *Rules and Regulators* (Clarendon Press, Oxford, 1997).

³² There is an empirical assumption in this paper that legal regulation that is highly complex, that confronts a lot of flux and that involves a lot of money tend to go together. There are I suspect good reasons in post-industrial information economies why this is so. However, it is of course not always so and a weakness of this essay is that it does not consider cases of stable complexity where not a lot of money is at stake, simple flux where a lot of money is at stake and so on. Of course, as discussed above, complexity, change and cash are readily conceived as continuous variables, so that: Certainty of Principles – Certainty of Rules = f (complexity + change + economic interests at stake).

more vague and less certain than rules.”³³ This empirical claim about how law works becomes a standard positivist normative proposition:

Since the law should strive to balance certainty and reliability against flexibility, it is on the whole wise legal policy to use rules as much as possible for regulating human behaviour because they are more certain than principles and lend themselves more easily to uniform and predictable application.³⁴

Is the empirical assumption about how the law works in practice that underpins this common normative position correct? My hypothesis is that it is correct with simple, stable patterns of action that do not involve high economic stakes, false with complex actions in changing environments where large economic interests are at stake – the staples of state regulatory practice and of private litigation in the appellate courts. All the great positivist jurists seem guilty of Robert Baldwin’s charge in his essay “Why Rules Don’t Work”.³⁵ This is that rules are drafted and argued about in legal texts without working through how compliance with them is secured in regulatory practice by police, administrative agencies, private litigants and even courts. When flux is great it can be obvious that radically abandoning the precision of rules can increase certainty. For example, Andrew Simpson makes the case that the regulation of transitional technologies like telecommunications requires “redefinable terms” in rules.³⁶ What a telephone means today may be something quite different tomorrow. In terms of the conclusion of this article, so long as the principles that underpin the redefining are clear, redefinable rules can regulate a transitional technology with more certainty than fixed rules.

Hart points out that rules have a core meaning and a penumbra where their meaning is more uncertain.³⁷ The more complex and changing the phenomenon being regulated, the wider that penumbra is likely to be; indeed in the most difficult contexts the penumbra of uncertainty swallows up the core creating large numbers of laws that are never enforced. A factor that does a great deal to drive uncertainty is that wealthy legal game players aim for the penumbra, play the game in ways that expand the grey area of the law. If they are repeat players in a particular regulation game, they may play for rules rather than for outcomes. The state sometimes does this as well – running cases they believe they will lose but which will deliver them

³³ Raz, above n 19 at 841.

³⁴ Id at 841.

³⁵ Robert Baldwin, ‘Why Rules Don’t Work’ (1990) 53(3) *Modern Law Review* 321.

³⁶ Andrew F Simpson, *Regulating Transition: Australian Telecommunications Industry Liberalization, 1989-1999* (2000) (unpublished PhD thesis, Faculty of Law, University of Sydney).

³⁷ Hart, above n 34.

a judicial interpretation of the law that will help them win future cases. One might say this game playing will not necessarily increase uncertainty. Whether it does that depends on whether the player who seeks to expand the uncertain penumbra of a rule succeeds or whether it is a player on the other side who is seeking to expand or defend the settled core meaning of the rule. This is true. But it is also true that the game itself iteratively settles and unsettles the law.

Of our three conditions, the economic stakes is the most important one because arguably, by playing the penumbra game, even laws that regulate simple, stable phenomena can be rendered complex. It might be said that one does not necessarily have to be a wealthy litigant fighting for high stakes to play this game. Little people who are extraordinarily highly motivated can play it as well. They can, but it is hard for them to sustain legal game playing for long without a lot of resources backing them.

Flux is a particularly important here. The penumbra of simple rules that regulate stable phenomena is small. It does exist as in Lon Fuller's example of a statue of a truck in a park where vehicles are prohibited.³⁸ The unforeseen rare event of the failure to encompass the truck statue in the rule will never cause great problems in regulating the static phenomenon of parks. But in dynamic domains like tax, uncommon things like concrete trucks can be rendered common by gameplaying investors seeking a tax advantage.³⁹

Positivism's penumbra problem is not the only game playing threat. Rules look more certain when they stand alone; uncertainty is created in the juxtaposition with other rules. With simple rules that regulate simple phenomena, it is easy to draft them so that there is no conflict between a rule that says A is proscribed and another that X is specifically allowable. The conflict arises when A and X are both correct descriptions of an action.⁴⁰ In complex terrain, however, economic and technological change can suddenly create new conflicts of this sort. It can be hard for the state to see this coming when there are well paid legal entrepreneurs on the lookout for opportunities to expand the penumbra of one rule to slightly overlap the

³⁸ Lon L. Fuller, 'Positivism and Fidelity to Law – A Reply to Professor Hart' (1958) 71 *Harvard Law Review* 630,663.

³⁹ "Uncommon transactions that are taxed inappropriately become common as taxpayers discover how to take advantage of them." David A. Weisbach, 'Formalism in Tax Law' (1999) 66 *University of Chicago Law Review* at 869.

⁴⁰ See McBarnet and Whelan's account of the use of "ex-files", express exemptions, in British accounting standards. Doreen McBarnet and Christopher Whelan, *Creative Accounting and the Cross-Eyed Javelin Thrower* (Wiley, Chichester, 1999) 103.

penumbra of another, creating compliant non-compliance.⁴¹ Indeed in areas like tax the discovery of such an overlap can be widely used for years without the state being aware of this. A firm can commit what the state would assume to be a breach of a criminal law secure in the knowledge that in the unlikely event that the well-concealed conduct is detected the company has a letter from a respected lawyer arguing that the conduct is legal by virtue of such an overlap. A court may find the lawyer's argument to be wrong, but so long as it is a "reasonably argued position"⁴² the letter advancing it will absolve the company of any criminal liability. This impunity combined with the huge benefits that can be secured motivates wealthy taxpayers in some cases to spend tens of millions of dollars on professional advice to such ends.⁴³ Moreover, with tax, legal game playing is not an activity of a small minority of the big players. An investment banker may not have been too wide of the mark in the late 1990s when he said to Joseph Bankman "There isn't a Fortune 500 Company that doesn't invest in these deals [the new US market in corporate tax shelters]".⁴⁴

This problem multiplies as the state enacts more and more rules to plug loopholes opened up by legal entrepreneurs. The thicket of rules we end up with becomes a set of sign-posts that show the legal entrepreneur precisely what they have to steer around to defeat the purposes of the law. Broad proscriptions against a phenomenon like insider trading can engender more certainty than a patchwork of specific rules that define A,B,C,D,E and F all as forms of insider trading. The rulish form of such an insider trading statute nurtures the plausibility of a legal argument that another form of insider trading – G – must be legal because the clear intent of the legislature

⁴¹ Id.

⁴² In the US, the test to secure such immunity is a "more-likely-than-not opinion". See Joseph Bankman, 'The New Corporate Tax Shelters Market' (21 June, 1999) 83 *Tax Notes* 1775.

⁴³ In the ASA case transaction costs were estimated at \$24.8 million (US), 27 per cent of the reported tax savings of \$93.5 million (US) (ASA Investering Partnership. Commissioner 76 T.C.M. (CCH) 325 (1998)). In cases of this ilk, the IRS has observed tax professionals to offer a choice of fees, either up-front payment of 25 per cent of taxes paid or a contingent fee of 50 per cent of taxes paid. PriceWaterhouseCoopers has been known to ask wealthy clients for a contingency fee of between 8 and 30 per cent of their tax savings. Deloitte and Touche have been revealed to be charging companies a contingency fee of 30 per cent of taxes saved for achieving a total wipeout of US state and federal taxes. See Department of the Treasury, *The Problem of Corporate Tax Shelters: Discussion, Analysis and Legislative Proposals* (Washington, DC, July 1999) 23. Janet Novack and Laura Saunders, 'The Hustling of X Rated Shelters' (14 December, 1998) *Forbes* 2.

⁴⁴ Bankman, above n 42. Bankman claims that Bear and Sterns & Co generated over \$100 million in fees for advice on lease-strip shelters in a matter of months.

was only to proscribe A to F, when in fact the legislature had never thought of G. Legal uncertainty arises from the fact that a thicket of rules engenders an argument of a form that some judges will buy and others will not. This produces one kind of very strong predictability in the law. If you know which judges you will and will not get, you can predict the outcome of your case with a high degree of certainty. In a multiple discriminant analysis of House of Lords decisions, well over 90 per cent of tax and criminal cases could be correctly predicted (in terms of whether the state or the other party – the taxpayer, the defendant – won) and more than 80 per cent of public, constitutional and civil cases could be correctly predicted in advance by knowing just one fact about that case.⁴⁵ That fact was which judges would sit on it. This kind of certainty is not much use to citizens, however, in pre-litigation decisions as to whether to back one view or another of what the law means.

A smorgasbord of rules engenders a cat and mouse legal drafting culture – of loophole closing and reopening by creative compliance.⁴⁶ Moreover it engenders a structurally inegalitarian form of uncertainty. The law thus engendered becomes so complex that little people who cannot afford sophisticated legal advice cannot understand it. In practice a particular law may be certain in the way lawyers apply it to them, but perceptually its complexity makes it uncertain to them as a guide to their actions, untutored as they are by legal advice. The rich, in contrast deploy legal entrepreneurship to make the law uncertain in practice. As citizens go about activities like paying taxes, creative compliance thus creates a law that is perceptually unclear to ordinary people, and therefore uncertain for them, and uncertain in practice for the rich who more clearly perceive and exploit this uncertainty.

Again we must remember that this structural fact of a formal lawyering of rules is not true of simple regulatory domains. There is no important sense in which traffic rules in a legal system are more incomprehensible to the poor than the rich or that they work in a way that enables the rich to make them more legally uncertain in practice in their application to the driving of the rich. The rich may be more able to bribe police and Magistrates in traffic cases, but this is not a matter of uncertainty in what the law says.

Legal entrepreneurship when economic stakes are high does not work simply by exploiting change and complexity that is inherent in post-industrial societies. It also works by contriving change and complexity. When the Australian government privatised Qantas on terms it thought were

⁴⁵ David Robertson, *Judicial Discretion in the House of Lords* (Clarendon Press, Oxford, 1998) 48-55.

⁴⁶ McBarnet and Whelan, above n 30 at 28.

generous for the new shareholders, it wanted only Australian citizens who had funded Qantas over the years through their taxes to gain the benefit. The Macquarie Bank, an Australian institution, responded by creating a new financial product called a QanMac to get around this rule; foreign QanMac owners secured identical functional economic benefits to Australian owners of Qantas shares. Financial engineering to create new products that have never been conceived by the law is a growing phenomenon of particular importance to corporations' law and tax compliance at the big end of town. "The same minds that figured out how to split a security into a multitude of different cash flows and contingent returns are now engineering products in which the tax benefits are split off from the underlying economics of a deal."⁴⁷ Financial engineering is just a newer modality of a more longstanding tradition of contrived complexity. Multinational corporations have long exploited their capacities to contrive complexity in their books, organizational complexity and jurisdictional complexity in order to escape liability even for comparatively simple criminal laws such as those against bribery.⁴⁸

Complexity in the books is used to enable the laundering of slush funds and deployment of a network of subsidiaries to contrive off balance sheet financing. Jurisdictional complexity can be exploited for example to shift losses to the jurisdiction where they will deliver a maximum tax deduction, profits to jurisdictions where gains are untaxed. Organizational complexity can take the form of the appointment of a "Vice-President Responsible for Going to Jail" to ensure there will be no corporate or CEO criminal liability, or more commonly it takes the form of a smokescreen of diffused accountability, where everyone can credibly blame someone else: there seem to be little bits of blame in many places without the possibility of aggregating this to a pattern that satisfies the rules of criminal responsibility.⁴⁹ Poor criminal defendants cannot contrive this kind of complexity into their affairs, which is why most inmates of our prisons are poor even though the evidence is clear that it is the rich who commit the criminal offences that cause greatest loss of property and injury to persons.⁵⁰ With respect to criminal law, this hardly plays out as a major source of uncertainty: a detected serious criminal offender of limited means is fairly certain to be convicted; major corporate criminals are almost certain never to be convicted partly because they are less likely to be detected and partly because of entrepreneurship in excuses and contrived complexity. Put another way; there is uncertainty that is structurally

⁴⁷ Novack and Saunders, above n 43.

⁴⁸ See John Braithwaite, *Corporate Crime in the Pharmaceutical Industry*, (Routledge, London, 1984).

⁴⁹ Id.

⁵⁰ See John Braithwaite and Philip Pettit, *Not Just Deserts: A Republican Theory of Criminal Justice* (Oxford University Press, Oxford, 1990) ch 9.

predictable by features of power in society rather than by features of the law.

Frederick Schauer argues: “In many cases, indeed in most cases, the result indicated by applying a rule will be the same as the result indicated by directly applying the rule’s background justifications...”⁵¹ What I am arguing is that this is true of the law of fraud applied to welfare cheats, false for the law of fraud applied to top management of large corporations. As the criminologists put it, the best way to rob a bank is to own it. It is one thing to show that a positivist like Schauer is wrong for large and important swathes of the law in action when he says that “Most commonly the application of a rule will be consistent with its justification”.⁵² It is quite another to show that you would get consistency by direct application of the justificatory principles. Indeed it seems on first consideration implausible that you would. If you cannot secure consistency with precise rules, why would you secure it with vague background principles?

Empirically it is quite difficult to explore the hypothesis that a rule of principles would be more consistent than a rule of rules. This is because of an empirical tendency for principles to go the way of rules.⁵³ Why this occurs is beyond the scope of this paper, but needless to say one reason is the very legal ideologies about the efficacy of tightly specified rules in assuring consistency that are being questioned in this essay. Moreover, it is hard to persuade legal officials in government to invest in research to test the validity of claims that they are already sure are true, especially when any demonstration that the law is administered inconsistently would result in public criticism of legal institutions. And these are only some of the obstacles.⁵⁴

⁵¹ Frederick Schauer, *Playing by the Rules: A Philosophical Examination of Rule-Based Decision-Making in Law and in Life* (Clarendon Press, Oxford, 1991) 100.

⁵² Id at 229.

⁵³ See McBarnet and Whelan, above n 30. See also Schauer, above n 1. Schauer gives an interesting account of convergence whereby principles tend to be filled out with rules (for example heuristic rules self-imposed by enforcement officials who seek to make their jobs less complicated by narrowing choice, or courts deciding specific cases with the effect that constraint by precedent becomes constraint by rule) and rules tend to go the way of principles (as when courts focus on the background justification of the rules or subject rules to a “reasonableness qualification”). Ultimately, my interest here is not to deny this, but to argue that what matters is how we set up the relationship between the bindingness of rules and the bindingness of principles.

⁵⁴ On the political difficulties with undertaking empirical research on the consistency of law see Valerie Braithwaite, John Braithwaite, Diane Gibson, Miriam Landau and Toni Makkai, *The Reliability and Validity of Nursing*

One research strategy would be to persuade a media company, say Court TV, to try a large number of cases twice with a panel of retired judges in a randomised controlled trial. Or retired judges could be hired to make real awards to real volunteer litigants in a laboratory setting. The objective would be to assess whether different judges deciding the same case using the rules of existing law would be statistically more likely to deliver consistent judgements than judges and/or juries instructed to rely only on a set of prescribed legal principles. A prediction of hypotheses 1 and 2 would be that rule-based decisions would prove more consistent with simple matters, principle-based decisions more consistent for complex matters. To add further research and viewer interest to a Court TV program, litigants and observers could be polled on whether they found the rule of principles or the rule of rules procedurally and substantively fairer, more respecting of their rights, win-win, win-lose, lose-lose, more dignified, less damaging to human relationships and so on. As our research group has found in Canberra, randomised controlled trials to courtroom adjudication of legal cases are very difficult to do, technically and politically. But it can be done and must be done if we want empirical answers to what are empirical questions.

An Empirical Case Study of the Certainty of Principles and of Rules

While we have no scientifically credible data to test our theory of the rule of principles versus a rule of rules at the level of judicial decision-making, we do have some rather revealing data at the level of decision making by state regulatory officials. Indeed it is this data that influenced me more than any other to specify the theory in this essay.

In 1988 the Australian federal government took over responsibility for nursing home regulation from state governments. Part of its reform agenda was to shift quality of care regulation from rules that mandated inputs and processes to “outcome standards.”⁵⁵ We have seen that many theorists do not distinguish standards from principles. Peter Drahos and I have distinguished them by conceiving of standards as defining measures of conduct: “Principles bring about mutual orientations between actors. Standards are norms that can be applied to measure their performance.”⁵⁶

Home Standards (Department of Health, Housing and Community Services, Canberra, 1991).

⁵⁵ See John Braithwaite, Toni Makkai, Valerie Braithwaite and Diane Gibson, *Raising the Standard: Resident-Centred Nursing Home Regulation in Australia*. (Department of Health, Housing and Community Services, Canberra, 1993).

⁵⁶ Braithwaite and Drahos, above n 1 at 19-20.

Standards vary in specificity. The Australian nursing home “standards” enacted in 1988 were by design low in specificity and therefore satisfy what I found above to be the core feature of Raz’s conception of principles: “Rules prescribe relatively specific acts; principles prescribe highly unspecific actions.”⁵⁷ The 31 outcome-oriented standards were negotiated during 1987 through the active collaboration of federal and state governments, industry, professional, union and consumer groups. The consensus standards that attracted the assent of all these groups were so broad that they were attacked by many as “motherhood statements”. I agreed with these critics at the time. Some were extremely broad outcomes: “The dignity of residents is respected by nursing home staff”. While it is arguable whether we should conceive of this as a principle or a standard, it is clearly not a Raz rule. In the course of the fieldwork in several countries observing nursing home inspections I became less cynical about “motherhood statements” when I contrasted how the Australian outcomes were enforced compared to more specified rules in other countries.⁵⁸ The Australian “homelike environment” standard was an example of what I had seen as a noble aspiration to require nursing homes not to be run like hospitals, not to be institutions. To my surprise we found that on the ground its enforcement had practical effects. If residents were not encouraged or helped to put family or religious pictures or art on their walls that was important to them, or were forbidden from bringing in some furniture or a carpet from their former home, this tended to be detected and resulted in the homelike environment standard being rated as not met. If residents wanted to have a pet in the home and this was ruled out of order by management, the inspectors challenged such a ruling and its reasonableness was discussed. The greatest contrast was some US states like Illinois, where I observed a lot of nursing home inspections, and where more specific rules oriented to the same objective were enforced. For example, one Illinois rule required counting the number of pictures on nursing home walls. This resulted in a common practice of nursing homes tearing one picture after another out of a few magazines and slapping them up with sticky tape above the residents’ beds prior to inspections.

Another example was the broad Australian standard: “Residents are enabled to participate in a wide range of activities appropriate to their interests and capacities.” In one nursing home I visited a resident invoked this standard to assert his right to visitations with a prostitute on a regular basis. This would have been an interesting one to specify as a rule. Some of the specification I saw on US inspections again had disturbing effects.

⁵⁷ Raz, above n 19.

⁵⁸ See John Braithwaite, ‘The Nursing Home Industry’ in A Reiss and M Tonry (eds), *Organizational Crime: Crime and Justice: A Review of Research* Vol 5 (1994) 11-54.

Counts were required to be recorded of the specific residents who attended each activity program. This resulted in sleeping residents in wheelchairs being wheeled in to the room where the activity was occurring; they were totally oblivious to the activity but were ticked off on the attendance book.

Most lawyers would have no trouble with the conclusion that the reliance on broad principles in the Australian regulatory model would result in wiser and more substantively effective regulation than the American rulebook. However, most would also have the intuition that the broad and vague Australian standards could not be enforced with any consistency. What does the evidence suggest?

In 1991 my research team undertook a reliability study of the Australian standards monitoring process.⁵⁹ We organized with the inspectorate for 50 nursing homes to receive two inspections at the same time. The inspectors made ratings of the compliance of the nursing home. These ratings agreed a minimum of 78 per cent of the time on the least reliable standard and as high as 98 per cent of the time on the most reliable standard. There was 88 per cent agreement across the 50 inspections on the homelike environment standard which increased to 94 per cent after inspectors were allowed to confer (so that one inspector could learn about things another inspector had seen that she had missed). The results were identical on the “activities” standard. After adding compliance scores across the 31 standards, we obtained inter-rater reliabilities that ranged between 0.93 and 0.96 depending on the stage of the regulatory process at which they were calculated.

Braithwaite and Braithwaite reviewed the results of four US inter-rater reliability studies of nursing home quality of care inspection.⁶⁰ All of them found considerably less reliability than the Australian results, one of them even finding no statistically significant association between the two sets of ratings. While there are many problems of comparability of the US and Australian studies, there can be no doubt that consistency in the legal outcome was massively less in the US.

The first reason the Australian standards were rated more consistently than the US standards is that there were so many of them in the US. At the time of our research, inspection teams were confronting more than a thousand standards in most US states, Australian teams 31. How do US

⁵⁹ John Braithwaite and Valerie Braithwaite, ‘The Politics of Legalism: Rules Versus Standards in Nursing Home Regulation’ (1995) 4 *Social and Legal Studies* 307. Valerie Braithwaite, John Braithwaite, Diane Gibson, Miriam Landau and Toni Makkai, *The Reliability and Validity of Nursing Home Standards* (Department of Health, Housing and Community Services, Canberra, 1991).

⁶⁰ *Id.*

inspectors cope with such a daunting task? The answer is that the human brain is incapable of coping. Some of the standards are completely forgotten, not suppressed by any malevolent or captured political motive, just plain forgotten. Such standards are never cited in the states where they are forgotten. Then there are those that become familiar by some accident of enforcement history that gave prominence to a particular standard in a particular state. Referring to state regulations, one midwestern inspector said during our fieldwork: “We use 10 per cent of them repeatedly. You get into the habit of citing the same ones. Even though possibly you could use others [for the same offence]. Most are never used.”

The professional background of the inspection team members is one important criterion that selects which standards will be attended to. Administrator: “If you’ve got a nurse, it will be nursing deficiencies in the survey report; if a pharmacist, you’ll get pharmacy deficiencies; a sanitarian, sanitary deficiencies; a lawyer, patient rights, etc.”. When inspectors have an impossible number of standards to check, our fieldwork showed how arbitrary factors cause particular standards to be checked in some homes, neglected in others, causing endemic inconsistency.

The classic process of writing more and more specific rules over time to cover newly discovered loopholes or apparent inconsistencies makes the body of rules as a package less capable of consistent assessment. The pursuit of consistency of parts causes the inconsistency of the whole. Our fieldwork suggested that US nursing home inspectors cope with the sheer cognitive demands of the enormous number of rules they are expected to be on top of by behaving rather like the way legal realists accuse judges of behaving.⁶¹ What the inspectors told us they did, and what we directly observed them to do, was operate from a gestalt of the prohibitions codified in the regulations – for example, that good infection control is required; that privacy must be protected; that good nursing practice should be followed. It is likely that professional training informs these gestalts more than the law. They then decide whether a citation ought to be written by deciding whether it offends against one of these gestalts. Then they search for the appropriate regulation under which to cite it. “What will we call it? How about 1220 A? What about 1220 B? Why don’t we use both of them?”

After explaining to a number of US inspectors this interpretation, based on our observations of how they coped, they agreed that this was basically how they did it. When we pointed out that the most troubling implication of this process from the point of view of reliability was that depending on how hard they searched through the standards, they might

⁶¹ That is deciding intuitively (in a manner not governed by rules) what is the right result and then scouring legal texts for the legal rules that will justify the intuition.

find one or two or three deficiencies to write out, one of them said, tellingly: "Or they might find none at all and have to mush it in." Decisions about how hard to search for multiple citations for essentially the same problem were driven by a "professional judgment" of "how serious overall their problems have been" or "how hard they've been trying". "You can write it out under [X] and create a repeat violation because they got a deficiency on [X] last time. Or you can write it out under [Y] so it's just an element, which has no real consequences. Or you can put it out under both [X] and [Y], putting out a whole standard."⁶² In this, nursing home inspectors perhaps cope in a way that is not radically different from the way the House of Lords copes with the galaxy of rules that constitute British law: "In many cases, and the Law Lords admit this readily enough, they work 'bottom-up', from a basic instinct that the plaintiff or the defendant ought to win to an argument that makes them a winner."⁶³ Both types of "basic instincts" are intuitions grounded in professional training more than personal values, though a critical difference is that with judges those professional intuitions are legal ones.

Hence, hand in hand with a paradox of consistency is a paradox of discretion. Lawmakers, in the misplaced belief that this narrows the discretion of inspectors, write more and more specific standards. The opposite is the truth: the larger the smorgasbord of standards, the greater the discretion of regulators to pick and choose an enforcement cocktail tailored to meet their own objective. A proliferation of more specific laws is a resource to expand discretion, not a limitation upon it.⁶⁴ The beauty of a small number of broad standards is therefore that one can design a regulatory process that ensures that the ticking of a met rating means that a proper process of information-gathering and inspection team deliberation has occurred on that standard. One accountability check in Australia was that whenever enforcement action was appealed, the team's worksheets listing all of the positives and negatives they found under each of the 31 standards were tabled before the Standards Review Panel.

An aspect of the US rules that is a particular source of uncertainty is the attachment of detailed protocols to US rules. Nursing home administrators love protocols. As one of them said, referring to protocols: "Give us the rules and we'll play the game". Imprecision and undefined evidence-gathering procedures make it hard for the efficient administrator to beat the system. In particular, there is no paper trail that can cover up a

⁶² This was a 1989 inspection which pre-dated abolition of the distinction between elements, standards and conditions of participation in Medicare.

⁶³ David Robertson, *Judicial Discretion in the House of Lords* (Clarendon Press, Oxford, 1998) 17.

⁶⁴ Robert Baldwin and Keith Hawkins, 'Discretionary Justice: Davis Reconsidered' (1984) *Winter Public Law* 570-599.

failure to grapple with the underlying justification for the rules: “You can achieve paper compliance without real compliance. You can fool most inspectors on most standards with paper compliance” (US interview). The source of unreliability then becomes the rare inspector who looks behind the paper trail to the quality of care that is actually given. Validity then becomes the major source of unreliability! One audit of the sources of inconsistency of New York inspections even put on the public record that one source was an inspector who instead of simply relying on “chart review” to assess decubitus ulcer treatment made the mistake of *observing* the care given:

The protocol states that only a chart review is necessary for this protocol, so the first cause for difference of opinion was a result of one surveyor doing more than he/she was instructed to do.⁶⁵

The other fundamental reason for the greater reliability of the ratings of the broad and vague Australian standards was that they establish a superior framework for regulatory deliberation than a large number of specific rules. We return to this matter later. First, I want to step back from this result that a set of standards (that are in fact principles in the terms defined here) result in more consistent regulatory judgments. We explore the theoretical possibility that a more judicious combination of rules and principles than we see in either Australian or US nursing home regulation might deliver superior consistency to that delivered by principles alone.

3. When the type of action to be regulated is complex, changing and involves large economic interests, principles or rules alone are less certain than a prudent mix of rules and principles.

Two possibilities here are: (a) rules that are legally binding backed by non-binding justificatory principles that assist the interpretation of those rules; and (b) non-binding rules that can be overridden by legally binding principles. When I say the rules are binding and the principles non-binding, I mean there is a presumption that the rules should be followed when there is a conflict between the rules and the principles. This presumption is reversed when it is the principles that are binding and the rules non-binding.

It is such a commonplace observation that rules cannot be interpreted without reference to background argument that I will not labour why this is so here.⁶⁶ It might follow that forging an explicit relationship between rules and principles that justify them will increase consistency. My focus will be on the comparative clarity of binding rules backed by non-binding principles versus binding principles backing non-binding rules. The first

⁶⁵ Office of Health Systems Management, *Evaluation of the New York State Residential Health Care Facility Quality Assurance System* (New York State Department of Health, 1985) 10.

⁶⁶ See, Black, above n 31, 129; Sunstein, above n 12.

option is close to Fred Schauer's suggestion of presumptive positivism.⁶⁷ Presumptive positivism means that most cases are decided by reference to the most locally applicable set of rules (which are presumptively binding). The rules have a priority, but not an absolute priority. If they produce a clearly unreasonable result, not merely a suboptimal result, when viewed from the perspective of the wider normative universe, then the rule can be abandoned in favour of some more profound principle. For example, presumptively a 100 kilometres per hour speed limit gets enforced, even in conditions when it is perfectly safe to drive at 110. However, the rule can be overturned if one is rushing to hospital with a dying person. Schauer suggests that presumptive positivism "may be the most accurate picture of the place of rules within many modern legal systems."⁶⁸

In the regulation literature a lot of the criticism of "Why Rules Don't Work" is a criticism of regulated actors being presumptive positivists. For example, the Kemeny Commission on the Three Mile Island nuclear accident diagnosed the problem to be that the nuclear plant operators had been educated by the regulatory system to be rule-following automatons.⁶⁹ They had come to rely heavily on the rules for guidance to the neglect of systemic understanding of the complex safety problem they were managing. They "gave insufficient emphasis to principles" according to the President's Commission on the Accident at Three Mile Island.⁷⁰ The shift away from presumptively positivist rule-following after Three Mile Island was associated with a dramatic improvement in nuclear safety as measured for example by the decline of the number of SCRAMS (automatic shut-downs of nuclear reactors) from 7 per unit in 1980 to 0.1 per unit in 1997.⁷¹ Jeremy Bentham's idea of "acoustic separation" of rules runs up against the same problem as presumptive positivism when confronted with complex regulatory challenges.⁷² Bentham suggested that there are different audiences for law: while judges should show interpretive flexibility to correct mistakes that would follow from slavish rule-following, what citizens should hear are clear general rules which must be followed. While this makes some sense with simple regulatory problems, with complex ones the Three Mile Island outcome looms large. We want the human objects of

⁶⁷ Schauer, above n 51.

⁶⁸ Id at 206.

⁶⁹ See Joseph Rees, *Hostages of Each Other: The Transformation of Nuclear Safety Since Three Mile Island* (University of Chicago Press, Chicago, 1994).

⁷⁰ JG Kemeny (Chairman), *The Need for Change: The Legacy of TMI. Report of the President's Commission on the Accident at Three Mile Island* 9-10 (GPO, 1979) 49.

⁷¹ Braithwaite and Drahos, above n 1 at 302.

⁷² Dan Cohen, 'Decision Rules and Conduct Rules: Acoustic Separation in the Criminal Law' (1984) 97 *Harvard Law Review* 625.

regulation as well as its subjects to grapple with the complexity of the phenomenon more completely than presumptive rule-following permits.

At the level of nursing home management we found many Benthamite acoustic separation theorists. Here is a quote from an Australian director of nursing that illustrates:

A checklist is my way of getting it done. We have to accept that we are dealing with girls who are rote learners. So it's the way I get them to learn. It's not checked off so they say, 'Oh, I haven't done it'.

In nursing home care as well, this rule-following mentality is a disaster for quality of care. As a result, staff who are more than just rote learners will show the personal integrity to rebel against it; they will get around the regulatory strictures of the checklist to respond particularistically to the manifest needs of residents in their care. At every level of a regulatory system inconsistent rule-following arises from intelligent people finding ways to honour the principles that justify the rules rather than the rules themselves. We observed a Chicago sanitarian point out during a nursing home exit conference that it is against the regulations to have a male and a female in adjoining rooms sharing the same toilet. The sanitarian points out that he understands that in this particular case neither resident is capable of using the toilet and that moving either of them would be upsetting to them. He points out that he is going to turn a blind eye to the rule for the sake of the residents, but he warns them that someone else from the Department could come along and cite them for this. In other words, he is pointing out that because there is such a mismatch between rule and outcome, he is giving an unreliable ruling. With Australian inspectors confronting such a predicament, there was no such unreliability. Since what was the best outcome for the residents was clear and since inspectors were instructed only to be concerned about outcomes, dialogue should quickly lead to a reliable result. In other words, greater inconsistency arises from the regime that attempts to achieve presumptive positivism because some will be persuaded to be positivists while others will work hard to corrupt positivism. This particular source of inconsistency does not arise when two law enforcers are trained to make all-things-considered judgements about underlying principles in uncontroversial cases like this.

Presumptive positivism may well be the legal theory that will result in greatest consistency with simple, stable problems where there are not large economic payoffs from game-playing with rules. But we have perhaps done enough to cast doubt on whether such a rule of rules would result in greater consistency than a rule of principles when the regulatory challenge is complex.

This leads us to consider the reverse option to normally binding rules and interpretive principles – binding principles backing non-binding rules. Considering this reversal is suggested by Julia Black’s account of the recent history of financial services regulation in the UK.⁷³ Black tells a story of self-regulation that was vague and principle-driven being replaced by state regulation motivated by a pursuit of certainty through “precise, specific rules”.⁷⁴ When some of the problems of a type we have already discussed arose, there was a substantial shift toward a rule of principles and a changing of the status of many of the detailed rules to non-binding guidance. This shift was far from total, however, and Black gives a more nuanced account of its content than I can capture here. Many participants in the British financial services market, on both sides of the fence, seem to have found some virtue in the shift away from binding rules and toward binding principles, while sustaining a mix of rules and principles.

So far in this essay we have not given any serious consideration to the advantages of rules. This is because it is not my objective to accomplish an overall assessment of the virtues and vices of rules, only the more limited objective of a theory of when rules are more and less consistently interpreted. But if the empirical explanation for inconsistent application of rules is that (some) enforcers subvert them because they see their vices as exceeding their virtues, the challenge of consistency becomes the challenge of institutionalising rules in such a way that rule-enforcers are less likely to find themselves in this predicament. Do non-binding rules backed by binding principles allow us to do this. Let us assume that all the virtues of rules articulated by Frederick Schauer are correct apart from his first “argument from reliance”, which is precisely what is in question in this article. His second is an “argument from efficiency”: following rules eliminates the necessity of making some kinds of investigations and calculations.⁷⁵ It may be that if rules are mere guidelines and we are accountable for honouring the principles that justify the rules we can acquire the wisdom to make contextual judgements as to when it is safe to get the efficiency benefits of following the rule and when it is not. This is what I, but not Schauer, take to be the virtue of a “rule of thumb”.

Schauer’s third case for rules is an “argument from risk aversion”: by insisting on rule following we might not achieve the optimal outcome but we eliminate the most foolish choices the all-things-considered decision maker might choose.⁷⁶ While Schauer’s analysis is insensitive to the limited predictability of the gravest and most complex risks of a risk society,⁷⁷ such

⁷³ Black, above n 31.

⁷⁴ Id at 216.

⁷⁵ Schauer, above n 51 at 146.

⁷⁶ Id at 146.

⁷⁷ Ulrich Beck, *Risk Society: Towards a New Modernity* (Sage, London, 1992).

as nuclear accidents or terrorism, there surely are some highly predictable serious risks. Might we not be able to highlight certain guidelines as ones we should be extremely resistant to flout because they protect against a palpably preventable risk? Surely we can avert the risks of missing unforeseen patterns of emerging risk by refusing to be rule-following automatons while understanding that there are some guidelines it would be extremely imprudent to ignore. Regulatory competence, whether for a judge or a factory inspector, may more than anything else be about getting wisdom about which rules should be stickier than others.

Next Schauer considers an “argument from stability”.⁷⁸ The argument for stability is a powerful one when rules regulate temptations of a rather stable character. But here we are considering the context of regulating complex economic and technological change where the argument from stability clearly has less appeal.

Schauer’s fifth virtue of rules is that they can be important tools for the allocation of power, guaranteeing in particular a separation of powers.⁷⁹ Yes we do need rules that program the polity: these kinds of decisions are better made by judges, other kinds by legislatures, others by the executive, others by a semi-autonomous central bank, others by the United Nations, others by the market, others by citizens without any interference from these other sources of power. But I have argued elsewhere that this ideal should not be one of rules of strict independence of these separated powers, but institutional arrangements that assure that all these separated powers can together check and balance abuses of power by other branches and that each branch has sufficient autonomy from the others to be able to exercise its power without being dominated by the other branches of power.⁸⁰ Some allocative rules with a lot of presumptive bite are doubtless needed here, but it is not at all clear that binding rules do more of the work of polities with meaningful separations of powers than non-binding conventions grounded in entrenched principles.

A final virtue of rules which Schauer identifies, but which lawyers often neglect, is that rules help us to escape from lose-lose outcomes in Prisoner’s Dilemma problems by communicating a salient win-win solution through the institutional salience of a rule.⁸¹ A non-binding rule might fulfil this useful purpose as well, or almost as well, as a presumptively positivist rule.

⁷⁸ Schauer, above n 51 at 155-158.

⁷⁹ Ibid at 158-162.

⁸⁰ John Braithwaite, ‘On Speaking Softly and Carrying Big Sticks: Neglected Dimensions of a Republican Separation of Powers’ (1997) 47 *University of Toronto Law Journal* 305-361.

⁸¹ Schauer, above n 51 at 162-165.

In different ways, using different reasoning, I suspect most people have a bottom line rather like mine in believing that we can achieve Schauer's advantages of rules without being a positivist or a presumptive positivist. More than that, with complex, ever-changing phenomena where the integrity of the rules are constantly under challenge from legal game playing by wealthy manipulators of the rules, there is not a lot of choice but to opt for wisdom in deciding which rules should be regarded as stickier than others. That said, we must call people to account for these judgments. This means enforceable principles that give reasons for why we should resist breaching the rules that should have the greatest stickiness and why even those sticky rules should be breached when doing so is imperative to safeguarding the principle that justifies them. Binding principles backing different degrees of stickiness in non-binding rules.

Saying this much still leaves a lot unanswered about how this would operate. Can we allow non-binding rules to create a safe harbour? This means that if you comply with certain non-binding rules you will be protected from enforcement action. Safe harbours involve large risks of failing to honour the principles in the law. Legal entrepreneurship in complex areas of law like tax is largely about steering clients into safe harbours (referred to as shelters in the tax context) that defeat the law's purposes. These risks are particularly acute with privately written non-binding rules pursuant to self-regulatory arrangements. To guard against these risks it is desirable that where profound public interests are at stake, privately written rules should be subject to public discussion and public ratification. Consider regulating the geologically complex and changing phenomenon of securing the roof of a mine. Roof falls are the major cause of death in modern mines.⁸² Here we actually need both the virtues of rules and the flexibility of rules that are particularized to a specific mine, and those rules need to be scrutinized by a safety regulatory agency and by union safety officials. Then, under an enforced self-regulation model,⁸³ the state should be able to prosecute a mining company criminally for failing to comply with its own rules. While there are no non-binding universal rules, there are principles that bind the firm to privately written particularistic rules. Again it follows from our analysis that complying with these rules should not be allowed to be a safe harbour. If the geology changes, it is the job of the firm to change its private rules rather than cling to them. In this changing context, it is the responsiveness of the rules, as opposed to an easy willingness to hide behind them, that is the best assurance of consistency by all participants in a regulatory community in honouring the principles that

⁸² John Braithwaite, *To Punish or Persuade: Enforcement of Coal Mine Safety* (State University New York Press, Albany, 1985).

⁸³ Ian Ayres and John Braithwaite, *Responsive Regulation: Transcending the Deregulation Debate*, (Oxford University Press, Oxford, 1992) ch 4.

justify the changing rules. So the empirical prediction is that if we send two inspectors in to independently assess whether the roof control of a coal mine honours overarching safety principles we are more likely to find agreement between them under such an enforced self-regulation dispensation. Agreement between the two inspectors will be less if there are no privately written and publicly ratified rules, only principles. And it will be less if universal presumptively positivist rules are in place. This kind of research can be done, as has been demonstrated with nursing home regulation.

4. Binding principles backing non-binding rules will be more certain still if they are embedded in institutions of regulatory conversation that foster shared sensibilities.

Again it is to Julia Black who we are indebted for the term regulatory conversation.⁸⁴ Black's main conclusion from the study of British financial services regulation was:

Normatively, building on an analysis of the nature of rules, it has been suggested that some of their inherent limitations may be ameliorated through the use of rule type, the development of interpretive communities, and the adoption of a conversational model of regulation.⁸⁵

For Black, certainty in the application of a rule can only arise from agreement about its terms within an interpretive community. Certainty does not flow so much from objective features of the clarity and precision of the words in rules, as lawyers sometimes assume, but from shared assumptions in a regulatory community about the interpreted shape of a rule. Indeed Black found that literalism in relation to a rule with sharp edges can be used as an interpretive strategy to defeat the purpose of a rule. For Black creative compliance is more than a failure to adopt a purposive approach: it is also "a refusal to 'read in' to the rule things which are suppressed by the generalizations or abstractions which the rule uses, and most significantly a refusal to recognize the tacit understandings on which the rule is based and on which it relies."⁸⁶ There is no escape of rule makers from such literalist game playing. Their best hedge against it according to Black are regulatory conversations which challenge failures to read in tacit understandings about

⁸⁴ Julia Black, 'Talking about Regulation' (1998) *Spring Public Law* 77.

⁸⁵ Black, above n 31 at 250.

⁸⁶ Julia Black, 'Managing Discretion' (paper to Australian Law Reform Commission Conference, *Penalties: Policy, Principles and Practice in Government Regulation*, Sydney, June 2001) 19.

the purpose of the rule, the state of the world and other matters that are relevant in the context.⁸⁷

Black's approach to regulatory conversations can help us make sense of the second major feature of why Australian nursing home regulation was more empirically certain than US regulation. When we spoke to senior regulatory bureaucrats in the United States and to social scientists who had been involved in the development and evaluation of nursing home inspection methodologies, a common type of comment was: "There are some things that the process cannot do reliably. So you don't do them. Examples are: 'Are the staff pleasant? Is the room tastefully decorated?'" The thought occurred to us that if the Hyatt hotel group adopted the view that décor and staff pleasantness were matters for which it could not set reliable standards (and therefore should not bother with), it would soon be losing money. In business, head offices effectively enforce all manner of "soft" standards on franchisees by adopting a qualitative approach to evaluation of performance where dialogue informs an evaluation. Surely one reason that American nursing homes are so cold, institutional, and inattentive to décor compared to say English nursing homes is precisely the attitude that such things are so subjective as to be beyond control.

The consistency of ratings of the "homelike environment" in Australia shows that the American posture on this matter is in error. A properly subjective approach on a standard such as this involves talking to residents about whether they feel free to put up personal things in an area

⁸⁷

The basic idea is that conversations are necessary for the interpretive conventions that make rules determinate. From their analysis of Saul Kripke's rule scepticism, Drahos and Parker conclude: "Conventions are pledges by interacting rule followers to extrapolate a rule in one unique way. So long as the conventions which are linked to a specific rule remain in place the rule can be said to be determinate." Peter Drahos and Stephen Parker, 'Rule Following, Rule Scepticism and Indeterminacy in Law: A Conventional Account' (1992) 5 *Ratio Juris* 117. Dimity Kingsford Smith invokes Stanley Fish's notion of an "interpretative community, whereby we understand the world, including texts such as the Corporations Law, by overlaying grids of intelligibility (or paradigms) through which we make sense of our experience. If these paradigms or structures are shared with other members of the interpretive community, then the meaning of the text being interpreted will be the same to all members. There is disagreement about whether these interpretive structures come from shared behaviour or activity, or from some sort of pre-interpretive understandings. But the essence of Fish's idea is that the closer the shared context of those using or interpreting a provision, the less indeterminate the meaning of that provision will be, and the more effective the community paradigm will have been in constraining meaning." Dimity Kingsford Smith, 'Interpreting the Corporations Law – Purpose, Practical Reasoning and the Public Interest' (1999) 21 *Sydney Law Review* 175.

they define as their private space, whether there are spaces in the facility that they feel are inviting and homelike places to chat with friends, whether they feel there are inviting garden areas they can use. This subjectivity often comes under attack in Australia. For example, managers of nursing home chains have complained to us that they have provided exactly the same food to two nursing homes; the team in one home gives them a “met” rating for the food and in the other home they get an “action required” rating. There is absolutely no inconsistency here if the residents at the two homes have different views about the food. Two teams might never agree on what is nice food, but we have found that they can agree, with high reliability, on whether the residents in a nursing home generally like the food they are getting. Reliability is accomplished by rejecting objectivity in favour of subjectivity.

Because consistency comes in this particular regulatory domain from being resident-centred, a path to improved consistency is better embedding the regulatory process in conversations with residents: learning how to draw out residents who have been intimidated into silence, how to capture moments of clarity of thought that normally confused residents experience, how to use third parties (roommates, relatives, group discussions with the Residents’ Committee) to draw out uncommunicative residents, translation support, and so on. In all regulatory domains there is a need to work at improving communication within the regulatory inspectorate, between the inspectorate and regulated actors and between both and the judiciary. There are many ways of institutionalising these possibilities through training, putting industry personnel on inspection teams, regular industry-wide roundtables, deliberative exit conferences where reasons for regulatory decisions are discussed, challenged and revised, and so on.

We have argued that a wise mix of rules and principles may be important to regulatory consistency. Nevertheless, just as Shearing and Ericson found police culture to be a story book rather than a rule book,⁸⁸ so we suspect regulatory cultures develop consistency more through stories that embody principles than through rules. Shearing has argued that stories constitute a sensibility from which action flows without reference to rules. They show participants in a regulatory culture how to “read”, via a “poetic apprehension” the layers of meaning in a story about principles that should guide them. A hotel chain does not secure quality décor through décor rules, but through stories and concrete examples of abominable and impeccable taste that nurtures the sensibilities central to this kind of private regulation.

Cross-national research from other domains of regulation beyond nursing homes also suggests that our conversational theory may be right

⁸⁸ Clifford D Shearing and Richard V Ericson, ‘Towards a Figurative Conception of Action’ (1991) 42 *British Journal of Sociology* 481-506.

concerning the importance of creating regulatory agency culture as a storybook compared to rulebooks as a route to certainty:

Edward Rubin reports that the German Bundesbank's regulations for assuring bank safety and soundness are bound in a pamphlet less than one hundred pages in length. The US Federal Reserve Board's operative statutes and regulations fill several thick binders. Officials hired by the Fed to work on bank regulation, aside from having a law degree or an adequate grade on a civil service examination, receive a few weeks of on-the-job training. The Bundesbank runs a three-year 'college' for its regulatory recruits. When regulatory officials are thoroughly trained professionals, dedicated to a career in the same regulatory program, Rubin notes, authorities can trust them to make programmatically sensible judgments and need not bind their discretion with detailed rules. Repeatedly, officials of multinational corporations whom we interviewed commented on frequency of turnover among the regulatory personnel they deal with in American agencies – which in turn led to variability in American regulators' level of technical knowledge when compared with their counterparts in Europe and Japan.⁸⁹

Studies of private law are as relevant here as public regulation. The historical and comparative insights in Hugh Collins' *Regulating Contracts* support an argument that contract law is more calculable when there is a move from formalism to rules and principles of contract embedded in the regulatory conversations that are the day to day stuff of business relations of contracting.⁹⁰ Collins' work cues us to the possibility that principles in the law might be interpreted with certainty by business actors who intuitively understand the business relations at issue, but with uncertainty by courts that do not grasp these intuitions.⁹¹ Moreover, he concludes that where alternative dispute resolution is conducted by arbitrators and mediators imbued with the custom and usage of the trade, it secures more consistency than formalistic courts.⁹² The remedy again might be the kind of improved conversation between business and the courts that was the accomplishment of Lord Mansfield.⁹³ Formalism can lock out regulatory learning, but there is no guarantee that more open textured principles will lock in regulatory

⁸⁹ Robert Kagan, 'The Consequences of Adversarial Legalism' in R A Kagan and L Axelrad (eds), *Regulatory Encounters: Multinational Corporations and American Adversarial Legalism* (University of California Press, Berkeley, 2000).

⁹⁰ Hugh Collins, *Regulating Contracts* (Oxford University Press, Oxford, 1999).

⁹¹ See also Anthony J Duggan, 'Is Equity Efficient?' (1997) 113 *Law Quarterly Review* 631.

⁹² Collins, above n 90, 187. Here Collins cites R L Bonn, 'The Predictability of Nonlegalistic Adjudication' (1972) 6 *Law and Society Review* 563.

⁹³ Collins, above n 90 at 188.

learning. A conversation between the justice of the people and the justice of the law is needed for that.

Principles may be more important than rules for engendering legal certainty because they are simpler, shorter, fewer and therefore can be better discussed from board meetings down to factory floor in drawing out the lessons from valued stories. The trouble with all this is that something has to be important enough to justify the investment in storytelling and other forms of deliberation that makes this work. Precise rules are regularly superior for the less important stuff of life, like telling us that we have to pay the parking meter between the hours of 8.30 am and 4 pm. The more complex and economically important matters that are our domain of interest here tend to be domains where participants in a regulatory community – firms, regulators, professionals, NGOs – are more likely to be willing to make a conversational investment. But often they will not for reason of politics or resources. Because this is often so, conversational regulation needs to be backed up with a default of punitive regulation against specific rules and general principles that close off the loopholes that can be woven through the rules.

Conclusions

This has been an attempt to develop a theory of legal certainty and to show that questions like whether presumptive positivism is a legal theory that should attract our allegiance depends on testing its empirical claims and assumptions about how rules work. The theory we have come to has three propositions:

1. When the type of action to be regulated is simple, stable and does not involve huge economic interests, rules tend to regulate with greater certainty than principles.
2. When the type of action to be regulated is complex, changing and involves large economic interests:
 - (a) principles tend to regulate with greater certainty than rules;
 - (b) binding principles backing non-binding rules tend to regulate with greater certainty than principles alone;
 - (c) binding principles backing non-binding rules are more certain still if they are embedded in institutions of regulatory conversation that foster shared sensibilities.

The empirical support for these propositions is extremely limited. Nevertheless, the propositions are advanced as more consistent with the little we know than the competing formulations considered. Even if the theory were right, given the nature of creative compliance, it would be naïve to think that the theory could be implemented in practice in a way that would not be at least partially corrupted by those with an interest in

contrived complexity. Paradoxically, it is lobbying for “business certainty”, for “telling business in advance what the rules are”, from the top end of town that persistently causes principles to go the way of rules.⁹⁴ And there are always positivist judges who will march beside the business lobbyists under the banner of certainty (in a way that in practice defeats certainty).⁹⁵

McBarnet and Whelan have even shown in British financial regulation how the loophole-closing principle of “substance over form” was allowed to be interpreted by legal entrepreneurs as a rigid rule so that the principle itself was used to create new avoidance devices.⁹⁶ It takes a judiciary and legislature more impervious to legal formalism than any have been in the past to hold the line on a principle like substance over form as an override of rigid rules (rather than a rigid rule in itself). Nevertheless, in the messy business of making law, holding the line on principles that override rigid rules may be achievable to greater and lesser degrees. To the extent that the propositions of the theory are approximated, then to that extent there may be more legal certainty.

Afterword: The Variable Contexts where Different Kinds of Rule-Principle Configurations Might Work

With the complex legal phenomena that preoccupy a large proportion of the time of our appellate courts, a big slice of state regulatory and self-regulatory expenditure, a large part of regulatory costs that burden the economy and some of the highest stakes for the community, the package of binding principles, non-binding rules and rich and plural regulatory deliberation may have some other advantages. Drawing on the work of

⁹⁴ See McBarnet and Whelan, above n 30.

⁹⁵ Robert Goodin interprets the recurrent tendency for principles to go the way of rules in a parsimonious way as a simple human predilection for wanting to exercise power that afflicts lawyers and business litigants like anyone else: “Power is the essence of politics, and the essence of power lies in restricting the choices available to others...A system of loose, principled law asks those in positions of power – legislators, judges, administrators, and so on – to pass up opportunities to exercise power. Thus it is not surprising that attempts at reintroducing loose, principled law prove abortive.” Goodin, above n 26, 71. (The implication of this essay is that while we can understand why they grab at power through rules in the way they do, in fact their grasp on it is frequently illusory.)

⁹⁶ McBarnet and Whelan, above n 30 at 866.

Robert Kagan and John Scholz⁹⁷ and Robert Baldwin,⁹⁸ Julia Black attempts to define which type of rule configuration will be more useful with regulatees who fail to comply for different reasons.⁹⁹ Some fail to comply because they are “political citizens” who have principled objections to the rules. Another group fail to comply because they are organizationally incompetent. Black argues that what we have defined here as principles are going to be most useful in helping these two groups to comply. Principles are more likely to engender the political and educative discussion needed to solve the non-compliance problem than “a large body of precise rules [that] can be difficult to absorb and alienating.”¹⁰⁰ For the organizationally incompetent these would need to be complemented with “user-friendly guidance manuals”.¹⁰¹ These are options enabled by the package of binding principles, non-binding rules and rich and plural regulatory deliberation.¹⁰²

For “irrational non-compliers” who resist the state telling them what to do even though it is rational to cooperate with the state, “Precise rules tend to be more effective as they enable the enforcer to show that the rule has been breached...”¹⁰³ The non-binding rules in the package advanced here certainly can be precise though their non-bindingness and the need to

⁹⁷ Robert Kagan and John Scholz, ‘The Criminology of the Corporation and Regulatory Enforcement Strategies’ in K Hawkins and J Thomas (eds), *Enforcing Regulation* (Kluwer-Nijhoff, Boston, 1984).

⁹⁸ Robert Baldwin, *Rules and Government* (Oxford University Press, Oxford, 1995).

⁹⁹ Black, above n 86 at 23-25.

¹⁰⁰ Id at 24.

¹⁰¹ Id at 24.

¹⁰² In these respects, Japanese regulation may closer approximate this principle-driven approach than Western regulation: “The book of effluent control regulations in Japan is ‘this thin’ a QUSA [multinational] environmental manager told us, holding two fingers an inch apart. The material she had to master in the United States, in contrast, filled a four foot bookshelf in her office... In Japan, we conclude, the regulatory regime appears to have gathered greater ‘normative gravity’, partly because Japanese officials view it as comprehensible, reasonable and predictable. This appears to facilitate the internalisation of regulatory norms by operating managers and workers. The fluctuating, polycentric character of the American regime, in contrast, appears to impair the law’s normative gravity.” Aoki Kazumasu, Lee Axelrad and Robert A Kagan, ‘Industrial Effluent Control in the United States and Japan’ in Kagan and Axelrad (eds), above n 89. Cross-national fieldwork such as in this project is an alternative method for testing the theory in this essay. This study concluded that the rule-based formalism of US environmental law generated less certainty than the more principle-based and dialogue-based Japanese approach.

¹⁰³ Id at 24.

underwrite enforcement with the demonstrated breach of a principle may be a disadvantage in dealing with this group.

“Amoral calculators” are for Black the hardest group to deal with:

neither detailed rules nor general rules are on their own likely to be successful in dealing with them. Detailed rules are vulnerable to strategies of ‘creative compliance’: compliance with the letter but not the spirit of the rule (McBarnet and Whelan 1991). General rules are vulnerable to challenges as to their interpretation and application. Amoral calculators are likely to contest the agency’s interpretation of the rule and assessment of compliance. The structure of appeal and review mechanisms then becomes highly relevant for determining who controls the interpretation of the rule, and thus the agency’s authority, *de facto* and *de lege*, for insisting that its interpretation is that which is ‘correct’.¹⁰⁴

While again the package advanced here has only a partial response to these challenges, it does not involve a total reliance on either the detailed or general. Tax enforcement in nations that have a General Anti-Avoidance Rule (GAAR), as the limited experience with them develops, may flesh out the possibilities here.¹⁰⁵ A GAAR is really a principle in the terms of this paper, which states that schemes are illegal when their dominant purpose is a tax advantage rather than a business purpose, even if the scheme “works” as a shelter from detailed tax rules.¹⁰⁶ Hence when the tax authority targets a shelter it can go after it first under specific rules in the law and if that fails

¹⁰⁴ Id at 24. The usage of “general rules” here might not be exactly in accord with the concept of principles in this essay. As Schauer points out, “the opposite of the specific is not so much the general as the *vague*. Not all general classes (or categories) are vague. The category ‘insects’, for example, is very large, including literally trillions of particular insects, but it is still reasonably specific, in the sense of *precise*.” Schauer, above n 20, 913. For the purposes of this essay, the opposite of specificity is non-specificity or vagueness, even though of course generality or broadness very often results in vagueness.

¹⁰⁵ These include Australia, Canada, Hong Kong and New Zealand in the common law world and a variety of civil law jurisdictions including Germany, France, Belgium, the Netherlands, Spain and Sweden. See Graeme S Cooper, ‘International Experience with General Anti-Avoidance Rules’ (2001) 54 *SMU Law Review* 83-130.

¹⁰⁶ The only other area of law where anything like a GAAR is common in developed economies is securities law. Weisbach illustrates with five “anti-abuse provisions” in US securities law. For example 17 CFR 230 Prelim Note 2 to Reg S (1998) (denies an exemption to foreign transactions to “any transaction or series of transactions that, although in technical compliance with these rules, is part of a plan or scheme to evade the registration provisions of the [Securities Act of 1933]”). David A Weisbach, ‘Formalism in Tax Law’ (1999) 66 *University of Chicago Law Review* at 884.

the tax authority can attack it under the GAAR.¹⁰⁷ Some tax experts think the GAAR something of an irrelevance because in jurisdictions that have it the courts rarely apply it, perhaps because they think it opens the door to giving too much discretion to the tax authority.¹⁰⁸ If that is so, the implication of the argument in this paper, if it is right, is that we need to persuade such courts that the reality is the opposite of their intuitions – a rule of rules that closes the judicial door on a GAAR actually reduces certainty and increases administrative discretion.¹⁰⁹ But it may be that for this to be true, judicial conversations need to be open to sharing the sensibilities of regulatory conversations between regulators and regulatees.¹¹⁰ Indeed in some way that does not threaten the separation of powers and judicial independence, judges need to become part of those conversations – Lord Mansfield’s project. To the extent that judiciaries prove incapable of this, then the case is strengthened for specialized courts such as the US Tax Court.

There are other solutions. Tax laws can be written by setting down binding principles, then detailed rules to illustrate how the principles should be applied to perhaps a dozen common concrete commercial

¹⁰⁷ At least this is the case under Australian law and administration.

¹⁰⁸ I am grateful for a discussion with Reuven Avi-Yonah on this question. The empirical observation that the GAAR is rarely used to decide cases in the courts in Australia and Canada is accurate, but it may not be correct that the reason for this is the courts fearing to give too much discretion to tax authorities. It may also be that the GAAR effectively deters schemes pre-litigation. See Cooper, above n 105 at 127. Rick Krever in commenting upon this paper made the point that GAARs are only used in practice “to deal with tax expenditures, not tax measures”. On this view, what is needed is not a GAAR, but a remedy to the problem of politicians wanting to introduce tax expenditures “without specifically spelling out who the intended beneficiaries are because it would be politically dangerous to explicitly identify the intended beneficiary group”.

¹⁰⁹ An additional reason for this with tax is that when one nation treats a particular kind of transaction more in terms of form and another state more in terms of substance, global firms can arbitrage substance and form by structuring transactions so that a part that gets a benefit under a form conception is located in the state that privileges form and another part that does better under a substance conception is channelled to a state that so treats it.

¹¹⁰ From my conversations with elite tax lawyers, there would be a lot of agreement with the following claim by David Weisbach, but not universal agreement: “I believe David Halperin’s claim that tax lawyers are sufficiently trained and share a sufficiently common understanding of the tax law to be able to determine which transactions anti-abuse rules target and which they do not.” Weisbach, above n 106 at 881 (citing David Halperin, ‘Are Anti-Abuse Rules Appropriate?’ (1995) 48 *Tax Law* 807, 809).

arrangements.¹¹¹ If there are 1000 rare ways of setting up the kinds of arrangements covered by the law, but only a dozen are used with any frequency then these are the 12 concrete arrangements that should be fleshed out into rules.¹¹² This means business is not left to flounder making sense of how to apply broad principles in their normal operations¹¹³ and are warned to be wary of tax advisers who counsel that the principles are so vague that a legal case can be made to justify almost anything in terms of them. At the end of the concrete rule specifications the law explains that these rules are defined in this way because they instantiate named principles. This is a further safeguard against lawyerly tendencies to privilege rules which can then be gamed. Another is an Acts Interpretation Act, such as those revised in New Zealand and Australia to complement their GAARs, that instruct courts to give effect to purposes of Acts¹¹⁴ or more specifically that defend the “the integrity of the tax system”.¹¹⁵ In spite of all of this, courts are still likely at times to indulge their proclivity to privilege the rules over the principles in the law say in response to a taxpayer whose advisers game one of the sets of rules that cover a concrete financial product or form of corporate structuring (X) into a minor variation (X!) that shelters income. Even then, there is an alternative to writing new rules to plug the loophole. It is for the legislature to enact a simple law that says the X! shelter violates named principles in the tax law and should be disallowed in future enforcement. Its effect is simply to strike down the court’s precedent in the X! case and to engage the judiciary in a

¹¹¹ See generally John Avery Jones, ‘Tax Law: Rules or Principles?’ (1996) 17 *Fiscal Studies* 63-72.

¹¹² Writing rules to cover the other 988 is precisely the drafting error we wish to avoid. Or, more precisely, we wish to avoid having to foresee them all and to avoid making the law iteratively more complex to cover them as they are used one after the other to game the law.

¹¹³ Another way to decide which arrangements should be defined by illustrative sets of rules would be to apply Louis Kaplow’s theory of when law should regulate through rules versus standards (principles in my usage). Simplifying, Kaplow contends that because rules have higher promulgation costs in deciding how to craft them ex ante, rules should only be written when the law will be applied frequently. Standards have lower promulgation costs than rules but higher application costs (costs in determining how they should apply to specific situations). Hence, standards are more economically efficient in application to arrangements that arise only rarely. Louis Kaplow, ‘Rules Versus Standards: An Economic Analysis’ (1992) 42 *Duke Law Journal* 557, 568-88.

¹¹⁴ Sections 15AA and 15AB Acts Interpretation Act, 1901 (Aust).

¹¹⁵ Section 6-6B New Zealand Tax Administration Act, 1994.

conversation with the legislature on the clarity of its intention to have a principle-driven tax law.¹¹⁶

While the strategy of attacking a shelter first in terms of rules and then in terms of a GAAR when this fails may have more general relevance beyond tax law, it is not likely to be appropriate for conduct where the community wishes to criminalize and imprison offenders. Where imprisonment is at risk people are entitled to know with some precision, and in advance, what puts them at risk of losing their liberty. In saying this, however, we have seen that we cannot assume that precise rules actually do protect us against arbitrary exercises of the most frightening state power when the economic phenomena at issue are complex.¹¹⁷ Fortunately, the criminal law is only occasionally the instrument the regulator who wishes to be effective in these areas wants to use. So the prescription here would be to never deprive a person of their liberty in circumstances where there is conflict between what is commended by a precise rule a citizen has relied upon and what is commended by the principle that justifies that rule.

We can make this move because most areas of law, including most criminal law enforcement, can be made to work perfectly well without resorting to locking people up. Some might want to make a similar move with human rights. Human rights, it might be argued, are so fundamental to liberty that they must be precisely formulated guarantees. But if the arguments in this article are correct, while relatively simple matters like a guarantee against detention without trial can and must be tightly specified rights, attempts to consistently guarantee a more complex right like a right to an education in a rulish fashion will not do well by such complex rights.

Julia Black's point about the control of the interpretation of principles is an important one. Pre-emptive self-regulatory rulemaking and pre-emptive private standard setting (for example through corporate capture of private standard-setting organizations) can deliver control of non-binding rules to amoral calculators. On the other hand, regulatory agency directives, rulings, clearances, waivers and legislatively mandated public ratification of private standard setting can hand that agenda to those who seek to restrain amoral calculators. The third way is for the state and concerned NGOs to persuade business interests who are not amoral calculators to prevent the amoral calculators from getting a competitive advantage over them by seizing the agenda for setting non-binding rules. One might suspect that the

¹¹⁶ I am indebted to conversations with Daniel Shaviro and Ernst Willheim for this thought.

¹¹⁷ Charles Black writes: "Some lawyers talk as though they thought maximum clarity always desirable even though they wouldn't have to probe very deeply to find that fraud, and fiduciary obligation, and undue influence, have been carefully isolated from exact definition..." Charles Lund Black Jr, 'Law as an Art', in *The Humane Imagination* (Oxbow Press, 1986).

conditions for the truth and falsity of the theory in this essay will turn greatly on whether control over non-binding rules works in the public interest, or rather in a plurality of interests, as opposed to being captured by self-interested legal game players. Practical issues like consumer group participation in organizations like the American National Standards Institute, the International Organization for Standardization and industry self-regulation schemes are of import here.

Baldwin's original thought in working through these types of non-compliers and the rule regime suggested by their compliance problem was that legislators should consider in advance whether regulatees were more likely to be political citizens, amoral calculators, organizationally incompetent or irrational non-compliers in a particular field. Then the form of rules could be crafted appropriately. Black's more ambitious thought is that "rather than opting for one rule type, rule makers should adopt a tiered approach to rule design in which rule types are combined in such a way that each tries to compensate for the limitations of the other".¹¹⁸ This essay can be read as an extremely modest addition to that ambitious agenda.¹¹⁹

¹¹⁸ Black, above n 31 at 24.

¹¹⁹ There are other agendas rather like this one. For example, Geoffrey Brennan has the agenda of defining regulatory contexts that are minimin and those that are maximax. Minimin institutional design seeks to eliminate the worst-case scenarios – to protect us from knaves as Hobbes would have it. This Hobbesian concern is also one of Schauer's in *Playing by the Rules*. Sometimes, in contrast, we want institutions that maximize our best shot (maximax) rather than protect us from the worst case. Universities and Olympic teams are examples, where we might want to maximize Nobel Prizes and gold medals. To some, it might not be clear that we would ever want regulatory institutions to be maximax. Christine Parker's work shows that indeed quite often we should because fields like protection of the environment are often pulled ahead by the environmental innovators more than pushed from below by regulators who knock out the worst practices of the worst firms. (Christine Parker, *The Open Corporation: Self-Regulation and Corporate Citizenship* (2002). For Parker, the first stage of internalising for maximax corporate citizenship is dialogue that leads to commitment to a set of principles. Clearly a legal rule of principles accommodates this whereas "[r]ules doom decision making to mediocrity by mandating the inaccessibility of excellence" (Sunstein, above n 12 at .177). In contrast, precise rules are more likely to be needed for the minimin regulatory objective. Rules, as Schauer and others explain, give up some possibility of excellence in exchange for guarding against the most dangerous forms of mediocrity. Again, it may be that the package of binding principles, non-binding rules and rich deliberation might accommodate maximax where it is needed and minimin where it is needed.