Institute for Austrian and International Tax Law

Welthandelsplatz 1, Building D3, 1020 Vienna
☏ +43/1/313 36/4890
www.wu.ac.at/taxlaw

WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS

JURISPRUDENTIAL PERSPECTIVES OF TAXATION LAW

CASES AND MATERIALS

Professor John Prebble QC

(Victoria University of Wellington)

Volume II

Master (Ausländisches Steuerrecht)
DIBT (Doctoral Program in International Business Taxation)

2022

Jurisprudential Perspectives of Taxation Law

Cases and Materials

Compiled and Edited by John Prebble

Volume I

1. **Introduction**

1.1. Harris, JW *Legal Philosophies* London Butterworths 1980 1 - 5. 1

1.2. Dias, RWM, "Introduction" from *Jurisprudence* (1985) 5th ed, Butterworths, London, 3 - 22. 6

1.3. Prebble, John "Why is Tax Law Incomprehensible?" [1994] British Tax Review 380-393. 26

1.4. White, Patricia D (ed) "Contents" and "Introduction" from Volume 1, *Tax Law* (1995) ix - xvii. 40

2. **Legal and economic concepts of income**

2.1. Prebble, John, *Asomatous Income*, (1995) Institute for Taxation and Business Law, Conference Paper Fourth Annual New Zealand-Pacific Tax and Business Law Conference, Tonga, 1995. 49

2.2. Thuronyi, Victor "The Concept of Income" (1990) 46 Tax Law Review 45. The following pages only: "Introduction" 45 - 46; "Nature of the income concept: the Haig-Simons tradition 48 - 53; "Specific problems in defining income" 64 - 92. 56

3. **Income redistribution: socialism and social justice**

3.1. Rawls, John "The Main Idea of the Theory of Justice" and other extracts from *A Theory of Justice* (Cambridge, Mass: The Belknap Press of Harvard University Press 1971) 11 – 303, excerpted in Smith, JC and David N Weisstub (eds) *The Western Idea of Law* (Canada: Butterworth & Co Ltd, 1983) 479 – 490. 93

3.2. Hayek, Friedrich A. von "The Atavism of Social Justice", in *Social Justice, Socialism, and Democracy: Three Australian Lectures*, Centre for Independent Studies, Turramurra, N.S.W. 1979. 3 – 15. Call # JC578 H417. 105

3.3. Hayek, Friedrich A. von "'Social' or Distributive Justice" Chapter 9 of *Law, Legislation and Liberty, Volume 2, The Mirage of Social Justice* London : Routledge and Keegan Paul, 1973, 62 – 106. VUW Call #JC585 H417 L. 112

4. **Ectopia**

4.1. Prebble, John "Ectopia, Tax Law, and International Taxation [1997] British Tax Review 383-403. 166

4.2. *Regent Oil Co Ltd v Strick (Inspector of Taxes)* [1966] AC 295 HL. Headnote 295 - 296 and judgment of Lord Reid 310 - 326 only. 187

5. **Fictions**

5.1. Dias, RWM, "Fiction" from *Jurisprudence* (1985) 5th ed, Butterworths, London, 318 - 319. 206

5.2. Prebble, John *Ectopia, the Root Cause of the Ineliminable Fictions of Income Tax Law*, (2006) Working Paper presented to a conference of the Western Regional Council of the Institute of Chartered Accountants of India, Mumbai, 30 November 2006. 208

5.3. Fuller, Lon L, extract from "What motives give rise to the legal fiction?" from *Legal Fictions* (1967) Stanford University Press, Stanford 56-63. …. 230

5.4. Priban, Jiri "Legalist Fictions and the Problem of Scientific Legitimation" (2003) 16 Ratio Juris 14 - 36. This is a difficult article. Class members should read it only if they have time. 238

6. **Form and substance**

6.1. Salmond, JW *Jurisprudence* 3rd ed London 1910. "Law logically subsequent to the administration of justice" 12 - 14; "Defects of the Law", 23 - 27 (especially formalism, p 25). 261

6.2. Isenbergh, Joseph "Musings on Form and Substance in Taxation" (1982) 49 Chicago Law Review 859, 863 - 883 only. 269

6.3. *BP Australia Ltd v FCT* [1966] AC 224, PC. Headnote 224 - 227 and judgment of Lord Pearce 260 - 274 only. 292

6.4. *Heather v P-E Consulting Group Ltd* (1972) 48 TC 293 CA. Headnote 293 - 294 and judgments 320 - 329 only. 311

7. **Particular issues in form and substance**

7.1. Prebble, John extracts from "Avoidance and other consequences of publishing Commissioner's interpretation guidelines" (2004) 19 Australian Tax Forum, 245 - 246. 323

7.2. *Ensign Tankers (Leasing) Ltd v Stokes (Inspector of Taxes)* [1992] 1 AC 655, HL, 655 - 668 (669 - 686 omitted). 339

7.3. Endicott, Timothy "Law is necessarily vague" (2001) 7 Legal Theory 379 - 385. 353

8. **The interface of form, substance, economic factors and legal factors**

8.1. *IRC v Wesleyan & General Assurance Society* (1948) 30 TC 11 - 27 CA & HL; [1948] 1 All ER 555 HL. 357

8.2. *Federal Commissioner of Taxation v Myer Emporium Ltd* (1987) 163 CLR 199 (HC, FC) (edited). 374

8.3. *Peterson v Commissioner of Inland Revenue* t[2006] 3 NZLR 433, [2005] HKPC 5 (PC), majority judgment and edited version of minority judgment.

.... 384

8.4. Prebble, John, "The *Peterson* Case and its Impact on the Rules in *BNZ Investments Ltd* and *Cecil Bros*" (2006) in Adrian Sawyer (ed) *Taxation Issues in the Twenty-First Centuy,* Centre for Commercial and Corporate Law (Inc) Christchurch. 409

9. **Positivism**

9.1. Austin, John, *Austin on Jurisprudence* ed Campbell, Vol 1 (London; John_Murray 1885) 86 – 227 excerpted in Smith, JC and David N Weisstub (eds) *The Western Idea of Law* (Canada: Butterworth & Co Ltd, 1983) 510 - 518. 425

9.2. Prebble, John *Advance Rulings on Tax Liability* (Wellington 1986) 48 - 49. 435

9.3. Kelsen, Hans "What is the Pure Theory of Law?" (1960) 34 Tulane Law Review 269 - 276. 437

9.4. Hart, H.L.A., "Kelsen Visited" (1963) 10 UCLA Law Review 709 - 728, 709 and 717 - 722 only. 445

Volume II

10. **Natural law**

10.1. Prebble, Rebecca, Traditional Naturalism and Neo-Naturalism (2003) unpublished paper. 452

10.2. Sabine, George H, *A History of Political Theory* (1937 New York, Holt Rinehart and Winston) 163 – 173, (On Cicero) excerpted in Smith, JC and David N Weisstub (eds) *The Western Idea of Law* (Canada: Butterworth & Co Ltd, 1983) 345 – 351. 465

10.3. Dias, R.W.M. "Natural Law", from *Jurisprudence* 5th ed, London, 1985, pages 78 - 85, Machiavelli to Democracy & Kant, 475 - 476. (Readers should focus on Locke, 81 - 83). 472

10.4. Epstein, Richard A "Taxation in a Lockean World" 4 Social Philosophy and Policy 49 - 74. 481

11. **Evasion, and the evasion/avoidance interface**

11.1. Prebble, John "Criminal law, tax evasion, shams, and tax avoidance: Part i - Tax evasion and general doctrines of criminal law" (1996) 2 NZ Journal of Taxation Law and Policy, 3-16. 507

11.2. Prebble, John "Criminal law, tax evasion, shams, and tax avoidance: Part i i - Criminal law consequences of categories of evasion and avoidance" (1996) 2 NZ Journal of Taxation Law and Policy, 59-74. 521

11.3. Honoré, Tony "The dependence of morality on law" (1993) 13 Oxford Journal of Legal Studies 1 - 17, pages 1 - 12 only. 537

11.4. Zoë Prebble and Prebble, John "The Morality of Tax Avoidance: Why the Difference between Avoidance and Evasion is Insufficient to Ground a Moral Distinction" Working Paper presented at an International

Conference on Legal Ethics: Professional Ethics and Personal Integrity, Auckland, 23-25 June 2006 549

12. **Statutory interpretation**

12.1. Unger, RM *Law in Modern Society* (1976 New York, The Free Press, 192 – 203, excerpted in Smith, JC and David N Weisstub (eds) *The Western Idea of Law* (Canada: Butterworth & Co Ltd, 1983) 617 – 623. ... 585

12.2. Prebble, John "Should tax legislation be written from a 'principles and purpose' point of view or a 'precise and detailed' point of view?" [1998] British Tax Review 112-123. Readers familiar with the author's ectopia thesis may prefer to begin at page 117. 592

12.3. *Duke of Westminster v IRC* [1936] AC 1 HL, only the following: headnote 1 - 2, Lord Tomlin 19 - 21, Lord Russell of Killowen 21 - 25. ... 604

12.4. *MacNiven (Inspector of Taxes) v Westmoreland Investments Ltd* (2001) 73 TC 1 HL. Summary pages 1 - 6 and judgment of Lord Hoffman pages 60 to 78. 613

12.5. *CIR v Mitsubishi Motors New Zealand Ltd* [1995] 3 NZLR 513 - 520 PC...638

13. **Autopoiesis and income tax law**

13.1. Teubner, Gunther "Introduction to Autopoietic Law" in Teubner, Gunther, *Autopoietic Law: a New Approach to Law and Society* (1987, de Gruyter, Berlin, New York), 1, 1 – 11 only. 646

13.2. Ladeur, Karl-Heinz "The Theory of Autopoiesis as an Approach to a Better Understanding of Postmodern Law" (1999) European University Institute Working Paper, Law No 99/3 3 - 44, 3 - 18 only. 657

13.3. Prebble, John "Autopoiesis of Income Tax Law: a Problem for Reform" (2004) working paper (fragment) draft and notes. 673

13.4. Printed earlier in the materials. *IRC v Wesleyan & General Assurance Society* (1948) 30 TC 24 CA &HL; [1948] 1 All ER 555 HL. ....

14. **Autopoiesis and anti-avoidance rules**

14.1. Francois Ost, "Between Order and Disorder: the Game of Law", in Teubner, Gunther, *Autopoietic Law: a New Approach to Law and Society* (1987, de Gruyter, Berlin, New York) 70, 70 – 83 only. 677

14.2. Hikaka, Geraldine and John Prebble, "Legal autopoiesis and general anti-avoidance rules" (2005) Seventeenth Australasian Tax Teachers' Association Conference, Wellington, January 2005. 691

15. **Coherence, analogy, and Kant's categorical imperative**

15.1. Printed earlier in the materials. Dias, R.W.M. "Natural Law", from *Jurisprudence*, extracted and printed earlier in this volume. Concentrate on Kant. ....

15.2. Bird, Graham "Kant, Immanuel" in Hondereich, Ted *The Oxford Companion to Philosophy* Oxford 1995 435, "The categorical imperative" 436 - 437. 727

15.3. Kress, Ken "Coherence" in Patterson, Dennis A Companion to Philosophy of Law and Legal Theory" (Oxford 1996) 533 - 552 728

15.4. *Commissioners of Inland Revenue v John Lewis Properties plc* [2003] Ch 513- 551, [2002] EWCA Civ 1869. 741

16. **The general anti-avoidance rule ("gaar"), the rule of law, morality, and other jurisprudential perspectives**

16.1. Prebble, Rebecca and John Prebble "Does the use of General Anti-Avoidance Rules to Combat Tax Avoidance Breach the Rule of Law?" Working Paper presented at the Australasian Tax Teachers' Association Conference, Brisbane, 22–24 January 2007. 780

16.2. Prebble, John "Ectopia, Formalism, and Anti-Avoidance Rules in Income Tax Law" (1994) in W. Krawietz, N. MacCormick, and G.H. von Wright (eds) Prescriptive Formality and Normative Rationality in Modern Legal Systems, Festschrift for Robert S. Summers, Duncker and Humblot, Berlin, 367-383. Readers familiar with the ectopia thesis may choose to scan pages 373 - 379 rather briefly. 819

16.3. Printed earlier in the materials. Prebble, Zoe, and John Prebble "The Morality of Tax Avoidance: Why the Difference between Avoidance and Evasion is Insufficient to Ground a Moral Distinction" Working Paper presented at an International Conference on Legal Ethics: Professional Ethics and Personal Integrity, Auckland, 23–25 June 2006. ....

COPYRIGHT WARNING NOTICE

This set of Student Notes may be used only for the University's educational purposes. It includes extracts of copyright works copied under copyright licences. You may not copy or distribute any part of these Student Notes to any other person. You may not make a further copy for any other purpose. Failure to comply with the terms of this warning may expose you to legal action for copyright infringement and disciplinary action by the University.

JURISPRUDENCE[1]

John Austin

The matter of jurisprudence is positive law : law, simply and strictly so called : or law set by political superiors to political inferiors. But positive law (or law, simply and strictly so called) is often confounded with objects to which it is related by resemblance, and with objects to which it is related in the way of analogy : with objects which are also signified, properly and improperly, by the large and vague expression *law*. To obviate the difficulties springing from that confusion, I begin my projected course with determining the province of jurisprudence, or with distinguishing the matter of jurisprudence from those various related objects : trying to define the subject of which I intend to treat, before I endeavour to analyze its numerous and complicated parts.

A law, in the most general and comprehensive acceptation in which the term, in its literal meaning, is employed, may be said to be a rule laid down for the guidance of an intelligent being by an intelligent being having power over him. Under this definition are concluded, and without impropriety, several species. It is necessary to define accurately the line of demarcation which separates these species from one another, as much mistiness and intricacy has been infused into the science of jurisprudence by their being confounded or not clearly distinguished. In the comprehensive sense above indicated, or in the largest meaning which it has, without extension by metaphor or analogy, the term law embraces the following objects : laws set by God to his human creatures, and laws set by men to men.

The whole or a portion of the laws set by God to men is frequently styled the law of nature, or natural law : being, in truth, the only natural law of which it is possible to speak without a metaphor, or without a blending of objects which ought to be distinguished broadly. But, rejecting the appellation law of nature as ambiguous and misleading, I name those laws or rules, as considered collectively or in a mass, the *divine law*, or the *law of God.*

Laws set by men to men are of two leading or principal classes : classes which are often blended, although they differ extremely ; and which, for that reason, should be severed precisely, and opposed distinctly and conspicuously.

Of the laws or rules set by men to men, some are established by *political* superiors, sovereign and subject : by persons exercising supreme and subordinate government, in independent nations, or independent political societies. The aggregate of the rules thus established, or some aggregate forming a portion of that aggregate, is the appropriate matter of jurisprudence, general or particular. To the aggregate of the rules thus established, or to some aggregate forming a por-

be particularly cogent in the case of "Utu" where it could be said that the investors did not merely indirectly procure the lender to make the loan by paying $y to the production company but the lender's ability to lend $y to the investors depended on the indirect receipt of $y from them.

51. The production company attributed the whole of the consideration of $x+y which it received from the investors as consideration for making the film and nothing as consideration for procuring the loan. Where, however, a single consideration is given for the supply of two or more goods or services the Commissioner is probably entitled even without section 99 to go behind the allocation agreed between the parties and allocate the consideration among the several goods or services for which it was paid on a proper basis. The Commissioner could argue that $x should be treated as paid to the production company as consideration for making the film and $y for procuring the loan. On this basis $y would not form part of the cost of acquiring a depreciating asset and would not qualify for the deduction claimed.

52. If such an argument were countered by the investors' the ignorance of the scope of the arrangements which had been made on their behalf (as to which their Lordships say nothing) the Commissioner could contend in the alternative that the arrangement by which the production company allocated the payment of $y as part of the consideration for acquiring the film instead of as consideration for the procurement of the loan (which affected the investors whether or not they were parties to it) could be counteracted under section 99.

53. The Commissioner, however, has never put any such case forward, it is not supported by the necessary evidence or findings, and it is contrary to a concession made before the Board. It cannot be said to be unanswerable; and it is not open to their Lordships to adopt it.

the parts of it which are explained least easily. Terms that are the largest and, therefore, the simplest of a series are without equivalent expressions into which we can resolve them concisely. And when we endeavour to define them, or to translate them into terms which we suppose are better understood, we are forced upon awkward and tedious circumlocutions.

If you express or intimate a wish that I shall do or forbear from some act, and if you will visit me with an evil in case I comply not with your wish, the expression or intimation of your wish is a command. A command is distinguished from other significations of desire, not by the style in which the desire is signified, but by the power and the purpose of the party commanding to inflict an evil or pain in case the desire be disregarded. If you cannot or will not harm me in case I comply not with your wish, the expression of your wish is not a command, although you utter your wish in imperative phrase. If you are able and willing to harm me in case I comply not with your wish, the expression of your wish amounts to a command, although you are prompted by a spirit of courtesy to utter it in the shape of a request. "*Preces erant, sed quibus contradici non posset.*" Such is the language of Tacitus, when speaking of a petition by the soldiery to a son and lieutenant of Vespasian.

A command, then, is a signification of desire. But a command is distinguished from other significations of desire by this peculiarity : that the party to whom it is directed is liable to evil from the other, in case he comply not with the desire.

Being liable to evil from you if I comply not with a wish which you signify, I am *bound* or *obliged* by your command, or I lie under a *duty* to obey it. If, in spite of that evil in prospect, I comply not with the wish which you signify, I am said to disobey your command, or to violate the duty which it imposes.

Command and duty are, therefore, correlative terms : the meaning denoted by each being implied or supposed by the other. Or (changing the expression) wherever a duty lies, a command has been signified ; and whenever a command is signified, a duty is imposed.

Concisely expressed, the meaning of the correlative expressions is this. He who will inflict an evil in case his desire be disregarded, utters a command by expressing or intimating his desire : He who is liable to the evil in case he disregard the desire, is bound or obliged by the command.

The evil which will probably be incurred in case a command be disobeyed or (to use an equivalent expression) in case a duty be broken is frequently called a *sanction*, or an *enforcement of obedience.* Or (varying the phrase) the command or the duty is said to be *sanctioned* or *enforced* by the chance of incurring the evil.

Considered as thus abstracted from the command and the duty which it enforces, the evil to be incurred by disobedience is frequently styled a *punishment.* But, as punishments, strictly so called, are only a class of sanctions, the term is too narrow to express the meaning adequately....

It appears, then, from what has been premised, that the ideas or notions comprehended by the term command are the following. 1. A wish or desire, conceived by a rational being, that another rational being shall do or forbear. 2. An evil to proceed from the former, and to be incurred by the latter, in case the latter comply not with the wish. 3. An expression or intimation of the wish by words or other signs.

tion of that aggregate, the term law, as used simply and strictly, is exclusively applied. But, as contra distinguished to *natural* law, or to the law of nature (meaning, by those expressions, the law of God), the aggregate of the rules, established by political superiors, is frequently styled *positive* law, or law existing by position. As contra distinguished to the rules which I style positive morality, and on which I shall touch immediately, the aggregate of the rules, established by political superiors, may also be marked commodiously with the name of *positive law.* For the sake, then, of getting a name brief and distinctive at once, and agreeable to frequent usage, I style that aggregate of rules, or any portion of that aggregate, positive law : though rules, which are not established by political superiors, are also positive, or exist by position, if they be rules or laws, in the proper signification of the term.

Though *some* of the laws or rules, which are set by men to men, are established by political superiors, others are not established by political superiors, or are *not* established by political superiors, in that capacity or character.

Closely analogous to human laws of this second class, are a set of objects frequently but improperly termed laws, being rules set and enforced by *mere opinion,* that is, by the opinions or sentiments held or felt by an indeterminate body of men in regard to human conduct. Instances of such a use of the term *law* are the expressions "The law of honour," "The law set by fashion," and rules of this species constitute much of what is usually termed "international law."

The aggregate of human laws properly so called belonging to the second of the classes above mentioned, with the aggregate of objects improperly but by close analogy termed laws, I place together in a common class, and denote them by the term *positive morality.* The name morality severs them from positive law, while the epithet positive disjoins them from the law of God. And to the end of obviating confusion, it is necessary or expedient that they *should* be disjoined from the latter by that distinguishing epithet. For the name *morality* (or *morals*), when standing unqualified or alone, denotes indifferently either of the following objects : namely, positive morality *as it is,* or without regard to its merits ; and positive morality *as it would be* if it conformed to the law of God, and were, therefore, deserving of approbation....

Having suggested the purpose of my attempt to determine the province of jurisprudence : to distinguish positive law, the appropriate matter of jurisprudence, from the various objects to which it is related by resemblance, and to which it is related, nearly or remotely, by a strong or slender analogy : I shall now state the essentials of a *law* or *rule* (taken with the largest signification which can be given to the term properly).

Every *law* or *rule* (taken with the largest signification which can be given to the term properly) is a *command.* Or, rather, laws or rules, properly so called, are a species of commands.

Now, since the term command comprises the term law, the first is the simpler as well as the larger of the two. But, simple as it is, it admits of explanation. And, since it is the key to the sciences of jurisprudence and morals, its meaning should be analyzed with precision.

Accordingly, I shall endeavour, in the first instance, to analyze the meaning of "command" an analysis which, I fear, will task the patience of my hearers, but which they will bear with cheerfulness, or, at least, with resignation, if they consider the difficulty of performing it. The elements of a science are precisely

It also appears from what has been premised, that *command, duty,* and *sanction* are inseparably connected terms : that each embraces the same ideas as the others, though each denotes those ideas in a peculiar order or series.

"A wish conceived by one, and expressed or intimated to another, with an evil to be inflicted and incurred in case the wish be disregarded," are signified directly and indirectly by each of the three expressions. Each is the name of the same complex notion.

But when I am talking directly of the expression or intimation of the wish, I employ the term command : The expression or intimation of the wish being presented prominently to my hearer ; whilst the evil to be incurred, with the chance of incurring it, are kept (if I may so express myself) in the background of my picture.

When I am talking directly of the chance of incurring the evil, or (changing the expression) of the liability or obnoxiousness to the evil, I employ the term *duty,* or the term *obligation* : The liability or obnoxiousness to the evil being put foremost, and the rest of the complex notion being signified implicitly.

When I am talking immediately of the evil itself, I employ the term *sanction,* or a term of the like import : The evil to be incurred being signified directly ; whilst the obnoxiousness to that evil, with the expression or intimation of the wish, are indicated indirectly or obliquely.

To those who are familiar with the language of logicians (language unrivalled for brevity, distinctness, and precision), I can express my meaning accurately in a breath. Each of the three terms *signifies* the same notion ; but each *denotes* a different part of that notion, and *connotes* the residue.

Commands are of two species. Some are *laws* or *rules.* The others have not acquired an appropriate name, nor does language afford an expression which will mark them briefly and precisely. I must, therefore, note them as well as I can by the ambiguous and inexpressive name of "*occasional* or *particular* commands."

The term laws or rules being not unfrequently applied to occasional or particular commands, it is hardly possible to describe a line of separation which shall consist in every respect with established forms of speech. But the distinction between laws and particular commands may, I think, be stated in the following manner.

By every command, the party to whom it is directed is obliged to do or to forbear.

Now where it obliges *generally* to acts or forbearances of a class, a command is a law or rule. But where it obliges to a specific act or forbearance, or to acts or forbearances which it determines *specifically* or *individually,* a command is occasional or particular. In other words, a class or description of acts is determined by a law or rule, and acts of that class or description are enjoined or forbidden generally. But where a command is occasional or particular, the act or acts, which the command enjoins or forbids, are assigned or determined by their specific or individual natures as well as by the class or description to which they belong.

The statement which I have given in abstract expressions I will now endeavour to illustrate by apt examples.

If you command your servant to go on a given errand, or not to leave your house on a given evening, or to rise at such an hour on such a morning, or to rise

at that hour during the next week or month, the command is occasional or particular. For the act or acts enjoined or forbidden are specially determined or assigned.

But if you command him simply to rise at that hour, or to rise at that hour always, or to rise at that hour till further orders, it may be said, with propriety, that you lay down a rule for the guidance of your servant's conduct. For no specific act is assigned by the command, but the command obliges him generally to acts of a determined class.

If a regiment be ordered to attack or defend a post, or to quell a riot, or to march from their present quarters, the command is occasional or particular. But an order to exercise daily till further orders shall be given would be called a general order, and might be called a rule.

If Parliament prohibited simply the exportation of corn, either for a given period or indefinitely, it would establish a law or rule : a kind or sort of acts being determined by the command, and acts of that kind or sort being generally forbidden. But an order issued by Parliament to meet an impending scarcity, and stopping the exportation of corn then shipped and in port, would not be a law or rule, though issued by the sovereign legislature. The order regarding exclusively a specified quantity of corn, the negative acts or forbearances, enjoined by the command, would be determined specifically or individually by the determinate nature of their subject.

As issued by a sovereign legislature, and as wearing the form of a law, the order which I have now imagined would probably be called a law. And hence the difficulty of drawing a distinct boundary between laws and occasional commands. [pp 86-93]

A law obliges generally the members of the given community, or a law obliges generally persons of a given class. A particular command obliges a single person, or persons whom it determines individually.

That laws and particular commands are not to be distinguished thus, will appear on a moment's reflection.

For, *first*, commands which oblige generally the members of the given community, or commands which oblige generally persons of given classes, are not always laws or rules.

Thus, in the case already supposed ; that in which the sovereign commands that all corn actually shipped for exportation be stopped and detained ; the command is obligatory upon the whole community, but as it obliges them only to a set of acts individually assigned, it is not a law. Again, suppose the sovereign to issue an order, enforced by penalties, for a general mourning, on occasion of a public calamity. Now, though it is addressed to the community at large, the order is scarcely a rule, in the usual acceptation of the term. For, though it obliges generally the members of the entire community, it obliges to acts which it assigns specifically, instead of obliging generally to acts or forbearances of a class. If the sovereign commanded that black should be the dress of his subjects, his command would amount to a law. But if he commanded them to wear it on a specified occasion, his command would be merely particular.

And, *secondly*, a command which obliges exclusively persons individually determined, may amount, notwithstanding, to a law or a rule.

For example, a father may set a *rule* to his child or children ; a guardian, to his ward ; a master, to his slave or servant. And certain of God's *laws* were as

binding on the first man, as they are binding at this hour on the millions who have sprung from his loins.

Most, indeed, of the laws which are established by political superiors, or most of the laws which are simply and strictly so called, oblige generally the members of the political community, or oblige generally persons of a class. To frame a system of duties for every individual of the community were simply impossible : and if it were possible, it were utterly useless. Most of the laws established by political superiors are, therefore, general in a twofold manner : as enjoining or forbidding generally acts of kinds or sorts ; and as binding the whole community or, at least, whole classes of its members.

But if we suppose that Parliament creates and grants an office, and that Parliament binds the grantee to services of a given description, we suppose a law established by political superiors, and yet exclusively binding a specified or determinate person.

Laws established by political superiors, and exclusively binding specified or determinate persons, are styled, in the language of the Roman jurists, *privilegia....*

It appears, from what has been premised, that a law, properly so called, may be defined in the following manner.

A law is a command which obliges a person or persons.

But, as contradistinguished or opposed to an occasional or particular command, a law is a command which obliges a person or persons, and obliges *generally* to acts or forbearances of a *class.*

In language more popular but less distinct and precise, a law is a command which obliges a person or persons to a *course* of conduct.

Laws and other commands are said to proceed from *superiors,* and to bind or oblige *inferiors.* I will, therefore, analyze the meaning of those correlative expressions ; and will try to strip them of a certain mystery, by which that simple meaning appears to be obscured.

Superiority is often synonymous with *precedence* or *excellence.* We talk of superiors in rank ; of superiors in wealth ; of superiors in virtue : comparing certain persons with certain other persons ; and meaning that the former precede or excel the latter in rank, in wealth, or in virtue.

But, taken with the meaning wherein I here understand it, the term superiority signifies might : the power of affecting others with evil or pain, and of forcing them, through fear of that evil, to fashion their conduct to one's wishes.

For example, God is emphatically the superior of Man. For his power of affecting us with pain, and of forcing us to comply with his will, is unbounded and resistless.

To a limited extent, the sovereign One or Number is the superior of the subject or citizen : the master, of the slave or servant : the father, of the child.

In short, whoever can *oblige* another to comply with his wishes, is the superior of that other, so far as the ability reaches : The party who is obnoxious to the impending evil, being, to that same extent, the inferior.

The might or superiority of God is simple or absolute. But in all or most cases of human superiority, the relation of superior and inferior, and the relation of inferior and superior, are reciprocal. Or (changing the expression) the party who is the superior as viewed from one aspect is the inferior as viewed from another.

For example, to an indefinite, though limited, extent the monarch is the superior of the governed : his power being commonly sufficient to enforce compliance with his will. But the governed, collectively or in mass, are also the superior of the monarch, who is checked in the abuse of his might by his fear of exciting their anger and of rousing to active resistance the might which slumbers in the multitude.

A member of a sovereign assembly is the superior of the judge : the judge being bound by the law which proceeds from that sovereign body. But, in his character of citizen or subject, he is the inferior of the judge : the judge being the minister of the law, and armed with the power of enforcing it.

It appears, then, that the term *superiority* (like the terms *duty* and *sanction*) is implied by the term command. For superiority is the power of enforcing compliance with a wish : and the expression or intimation of a wish, with the power and the purpose of enforcing it, are the constituent elements of a command.

"That laws emanate from superiors" is, therefore, an identical proposition. For the meaning which it affects to impart is contained in its subject.

If I mark the peculiar source of a given law, or if I mark the peculiar source of laws of a given class, it is possible that I am saying something which may instruct the hearer. But to affirm of laws universally "that they flow from superiors," or to affirm of laws universally "that inferiors are bound to obey them," is the merest tautology and trifling.

Like most of the leading terms in the sciences of jurisprudence and morals, the term *laws* is extremely ambiguous. Taken with the largest signification which can be given to the term properly, laws are a species of commands. But the term is improperly applied to various objects which have nothing of the imperative character : to objects which are not commands and which, therefore, are not laws, properly so called... [pp. 95-98]

The science of jurisprudence (or, simply and briefly, jurisprudence) is concerned with positive laws, or with laws strictly so called, as considered without regard to their goodness or badness.

Positive morality, as considered without regard to its goodness or badness, might be the subject of a science closely analogous to jurisprudence . I say "might be" since it is only in one of its branches (namely, the law of nations or international law), that positive morality, as considered without regard to its goodness or badness, has been treated by writers in a scientific or systematic manner. For the science of positive morality, as considered without regard to its goodness or badness, current or established language will hardly afford us a name. The name morals, or science of morals, would denote it ambiguously : the name morals, or science of morals, being commonly applied (as I shall show immediately) to a department of ethics or deontology. But, since the science of jurisprudence is not unfrequently styled "the science of positive law," the science in question might be styled analogically "the science of positive morality...."

The science of ethics (or, in the language of Mr. Bentham, the science of deontology) may be defined in the following manner. It affects to determine the test of positive law and morality, or it affects to determine the principles whereon they must be fashioned in order that they may merit approbation. In other words, it affects to expound them as they should be ; or it affects to expound them as they ought to be ; or it affects to expound them as they would be

if they were good or worthy of praise; or it affects to expound them as they would be if they conformed to an assumed measure.

The science of ethics (or, simply and briefly, ethics) consists of two departments: one relating specially to positive law, the other relating specially to positive morality. The department which relates specially to positive law, is commonly styled the science of legislation, or, simply and briefly, legislation. The department which relates specially to positive morality is commonly styled the science of morals, or, simply and briefly, morals.... [pp. 172-173]

The superiority which is styled sovereignty, and the independent political society which sovereignty implies, is distinguished from other superiority, and from other society, by the following marks or characters. 1. The *bulk* of the given society are in a *habit* of obedience or submission to a *determinate* and *common* superior: let that common superior be a certain individual person, or a certain body or aggregate of individual persons. 2. That certain individual, or that certain body of individuals, is *not* in a habit of obedience to a determinate human superior. Laws (improperly so called) which opinion sets or imposes may permanently affect the conduct of that certain individual or body. To express or tacit commands of other determinate parties, that certain individual or body may yield occasional submission. But there is no determinate person, or determinate aggregate of persons, to whose commands, express or tacit, that certain individual or body renders habitual obedience.

Or the notions of sovereignty and independent political society may be expressed concisely thus. If a determinate human superior, not in a habit of obedience to a like superior, receive habitual obedience from the bulk of a given society, that determinate superior is sovereign in that society, and the society (including the superior) is a society political and independent.

To that determinate superior, the other members of the society are *subject*: or on that determinate superior, the other members of the society are *dependent*. The position of its other members towards that determinate superior, is *a state of subjection*, or *a state of dependence*. The mutual relation which subsists between that superior and them, may be styled *the relation of sovereign and subject*, or *the relation of sovereignty and subjection*.

Hence it follows, that it is only through an ellipsis, or an abridged form of expression, that the *society* is styled *independent*. The party truly independent (independent, that is to say, of a determinate human superior), is not the society, but the sovereign portion of the society: that certain member of the society, or that certain body of its members, to whose commands, expressed or intimated, the generality or bulk of its members render habitual obedience. Upon that certain person, or certain body of persons, the other members of the society are *dependent*: or to that certain person, or certain body of persons, the other members of the society are *subject*. By "an independent political society," or "an independent and sovereign nation," we mean a political society consisting of a sovereign and subjects, as opposed to a political society which is merely subordinate: that is to say, which is merely a limb or member of another political society, and which therefore consists entirely of persons in a state of subjection.

In order that a given society may form a society political and independent, the two distinguishing marks which I have mentioned above must unite. The generality of the given society must be in the habit of obedience to a determinate and common superior: whilst that determinate person, or determinate body of

persons must not be habitually obedient to a determinate person or body. It is the union of that positive, with this negative mark, which renders that certain superior sovereign or supreme, and which renders that given society (including that certain superior) a society political and independent.

To show that the union of those marks renders a given society a society political and independent, I call your attention to the following positions and examples.

1. In order that a given society may form a society political, the generality or bulk of its members must be in a *habit* of obedience to a determinate and common superior.

In case the generality of its members obey a determinate superior, but the obedience be rare or transient and not habitual or permanent, the relation of sovereignty and subjection is not created thereby between that certain superior and the members of that given society. In other words, that determinate superior and the members of that given society do not become thereby an independent political society. Whether that given society be political and independent or not, it is not an independent political society whereof that certain superior is the sovereign portion.... [pp. 220-222]

Society formed by the intercourse of independent political societies is the province of international law, or of the law obtaining between nations. For (adopting a current expression) international law, or the law obtaining between nations, is conversant about the conduct of independent political societies considered as entire communities : *circa negotia et causas gentium integrarum.* Speaking with greater precision, international law, or the law obtaining between nations, regards the conduct of sovereigns considered as related to one another.

And hence it inevitably follows, that the law obtaining between nations is not positive law : for every positive law is set by a given sovereign to a person or persons in a state of subjection to its author. As I have already intimated, the law obtaining between nations is law (improperly so called) set by general opinion. The duties which it imposes are enforced by moral sanctions : by fear on the part of nations, or by fear on the part of sovereigns, of provoking general hostility, and incurring its probable evils, in case they shall violate maxims generally received and respected....

The definition of the abstract term *independent political society* (including the definition of the correlative term sovereignty) cannot be rendered in expressions of perfectly precise import, and is therefore a fallible test of specific or particular cases. The least imperfect definition which the abstract term will take would hardly enable us to fix the class of every possible society. It would hardly enable us to determine of every *independent* society, whether it were *political* or *natural.* It would hardly enable us to determine of every *political* society, whether it were *independent* or *subordinate.*

In order that a given society may form a society political and independent, the positive and negative marks which I have mentioned above must unite. The *generality* or *bulk* of its members must be in a *habit* of obedience to a *certain* and *common* superior : whilst that certain person, or certain body of persons, must *not* be habitually obedient to a certain person or body. [pp. 225-227]

ing from the Internal Revenue Service. Certainly, this practice is almost invariable where a prospectus relates to any kind of tax shelter. One sometimes sees a prospectus containing an adverse ruling with a statement by the promoter explaining why he disagrees with it.

These sorts of developments are sometimes thought to be a bad thing and to constitute a reason why a rulings proceess should not be established. That view does not seem to be shared by specialist tax practitioners. Moreover, it is based on a distinction that is apparent rather than real. An experienced and knowledgeable tax adviser can always venture an opinion on difficult areas of law. However, some areas are more difficult than others; consequently opinions must be proffered with varying degrees of certainty. Also, something that might seem to be certain to a metropolitan accountant engaged exclusively on corporate tax matters may not look so obvious to a provincial solicitor who is entirely competent but who is accustomed to handling a wider range of work. Thus, the difference between a problem of statutory interpretation that is virtually intractable and a problem about which one can give a confident opinion is a question of degree rather than of kind. If one accepts that law, and in particular tax law, should be certain in its application then the Canadian practice of applying for rulings to confirm opinions of tax advisers is not at all unreasonable.

§4.6 United States of America

In the United States there are a number of procedures which fall broadly under the heading of rulings.[9] First, there are 'letter rulings', which comprise written statements issued to a taxpayer by the national office of the Internal Revenue Service applying the tax laws to a specific state of facts. Secondly, there are 'revenue rulings' which are official interpretations published by the IRS for the information and guidance of taxpayers, IRS officials, and professional advisers. Thirdly, there are 'closing agreements' which consist of formal agreements between the IRS and taxpayers as to the tax consequences of proposed transactions.

Closing agreements have a statutory base, but letter and revenue rulings are matters of administrative practice. The reason for this difference is historical.[10] When the Revenue Act was enacted in 1913 the IRS answered all requests for rulings on both proposed and completed tansactions, although it did not consider itself to be bound by them. In 1919, after more complex legislation had been enacted, the IRS found that it was unable to cope with the volume of requests for rulings. Accordingly, in that year it was announced that rulings would be given only in relation to completed transactions and in certain specific cases where the Act required that a ruling had to be obtained before a transaction was completed.

The limited scope of the rulings procedure proved unpopular. Consequently, in 1938 legislative provision was made for 'closing agreements', whereby

the IRS and taxpayers could enter into a formal agreement as to the tax consequences of proposed transactions.[11] However, the closing agreement procedure was both slow and complex and the IRS began issuing letter rulings to meet the flood of requests for closing agreements. That state of affairs continued until 1953 when the IRS formally announced the existence of the letter rulings programme.[12]

Salient features of the United States of America's letter ruling procedure are:[13]

- As letter rulings were instituted by administrative action it follows that strictly speaking they are not binding on the authorities and can be ignored by the IRS. In practice, rulings are revoked where they are incompatible with a decision of the United States Supreme Court, where the law on which they are based is changed by legislation, and where there has been a material omission of fact in the original application. Generally, except in this last case, rulings are revoked prospectively rather than retroactively.[14]
- There is no appeal from an unfavourable ruling except where a ruling is required by the Internal Revenue Code 1954.
- Rulings are not issued where the IRS's position is uncertain as a result of an adverse judicial decision and the IRS has not decided whether that decision should be followed; where the transaction in question lacks a bona fide purpose or has as its principal object the reduction of tax; where the request contains alternatives; and where the request is in relation to a hypothetical situation.
- Requests for rulings are made to the National Office of the IRS. They must be in writing and must contain a complete statement of facts relating to the transaction along with the names and taxpayer identification numbers of all interested parties.
- The IRS does not charge fees for letter rulings.

9 See generally B I Bittker *Federal Income Taxation* (5 ed., Little Brown & Co., Boston, 1980) Vol 4, 34-36.

10 For an outline of the history of the United States of America's rulings programme see S.M. Goodman 'Notes — The Availability and Reviewability of Rulings of the Internal Revenue Service' (1964) 113 *U Penn L R* 81; M Rogovin 'The Four R's: Regulations, Rulings, Reliance and Retroactivity' (1965) 43 *Taxes* 756.

11 Closing agreements are currently made under Section 7121 of the Internal Revenue Code 1954.

12 *Rev Rul* 10, 1953-1 *Cum Bull* 488.

13 Details of the current procedure are contained in *Rev Proc* 80-20, IRB 1980 - 26, 7.

14 Section 17.05 *Rev Proc* 80-20, IRB 1980 - 26, 7; L W Kasischke 'New Procedures to obtain an IRS Ruling' (1980) 11 *Tax Adviser* 580, 587.

WHAT IS THE PURE THEORY OF LAW?

HANS KELSEN†

The Pure Theory of Law is, as its name indicates, a theory of law. The way in which a theory is elaborated is determined by its object. In order to apprehend the peculiarity of a theory of law, we must know the nature of its object; we must, first of all, answer the question as to what is law.

Although the theory of law — or, as it is usually called, jurisprudence — is one of the oldest sciences, there is no generally accepted definition of the concept of law. There are two different views concerning this object. According to the one, law is a *fact*, a definite behavior of men, which takes place in time and space and can be perceived by our senses. Facts are the object of the natural sciences: physics, chemistry, biology, psychology, sociology. Hence, according to this view of the law, jurisprudence does not essentially differ from these natural sciences. Just as these sciences, jurisprudence describes its object in statements to the effect that something is or is not, that is to say: in *is*-statements. According to the other view, law is not a fact, but a *norm*. A norm is a rule whose meaning is that something *ought* to be or to be done, even if actually it is not, or is not done. A norm has the character of a command or prescription and is usually expressed linguistically in an imperative, as, *e.g.*, the Ten Commandments God issued on Mount Sinai: "Honor your father and mother," "You shall not kill," and so on. A norm may have not only the character of a command but also the character of an authorization; by a norm a person may confer upon another person the power or capacity of issuing commands. God authorized Moses to issue commands to the Jewish people; God conferred upon him the authority of a legislator. The constitution of a state authorizes a certain individual or a body of individuals to issue statutes — general norms — and statutes authorize courts and administrative organs to issue individual norms — judicial decisions and administrative commands. Finally, a norm may have the character of a permission, that is to say, by a norm a person may be allowed to do something which without that permission is forbidden. For example, a general norm forbids killing, but by a special norm, restricting the first one, it is permitted as self-defense. These are the three functions of a norm: command or prescription; authorization; permission. If we say that a norm is a rule whose meaning is that something *ought* to be done, the term "ought" covers the meaning of all the three functions; it indicates the normative function.

The specific meaning of the statement that something *ought* to be or to be done can be determined only by referring to the differ-

†Professor Emeritus, University of California, Berkeley.

ence which exists between this statement and the statement that something *is* or is done. Of this difference we are immediately and directly aware. The logical dualism of the "ought" and the "is" implies the impossibility of inferring from the statement that something ought to be or to be done, the statement that something is or is done, and vice versa.

We are aware of the "ought" as of something different from the "is" if the former is the meaning of an act of an individual intentionally directed at the behavior of another individual. If, for instance, *A* commands *B* to do something, we describe the *act* by the statement that *A* wills that *B* do something. This is an *is*-statement. But the *meaning* of the act can be described only by the statement that *B* *ought* to do something, not by an *is*-statement such as that *B* does or will do what *A* commands. That *B* ought to do something is the *subjective* meaning of the act of command; the meaning the act has from the point of view of the commanding individual. But it is not necessarily also the *objective* meaning of the act of command, that is the meaning the act has from the point of view of the addressee and of a third person not concerned. That *B* ought to do something is considered to be also the objective meaning of the act of command if this act is authorized. Then, its meaning is called a *norm*. To the question as to the difference between the subjective and objective meaning of acts prescribing a certain behavior, we shall return later.

According to the Pure Theory of Law, law is norm, or, more exactly, a set of norms, a normative order. Since a normative order is the object of jurisprudence, and the meaning of norms is that something ought to be done, that men ought to behave in a certain way, jurisprudence can describe its object not as natural sciences describe their object — in *is*-statements — but only in *ought*-statements. To the question as to what is the law in a certain matter, as for example, with respect to theft or murder, the answer is not that if a man commits theft or murder he is or will be punished, but that he ought to be punished because the law exists — and that means: the law is valid — even when in a concrete case a thief or murderer actually is not punished, because, for example, he has escaped punishment. If in a textbook of the criminal law of California a statement were made that a man who commits theft or murder is or will be punished, such a statement would be false because, unfortunately, there are in California some cases in which — exceptionally — a murderer is not punished. But the law is that a murderer in all cases ought to be punished.

When the legal authority uses in the norm issued by it the term "ought" — for example, a thief ought to be punished, a civil execution ought to be directed against the property of a debtor who does

not repay his debt — this term has a *prescriptive* meaning. As a prescription or command, a norm is neither true nor false, but valid or not valid. However, the same term has a *descriptive* meaning when it is used by legal science in a statement affirming the validity of a legal norm. Such a statement may be true or false. Only the legal authority can *pre*scribe; legal science can only *de*scribe what the legal authority prescribes.

Although the principles of logic, such as the law of contradiction and the rules of inference, apply only to statements which can be true or false, they are indirectly applicable also to legal norms, insofar as statements about norms, statements affirming the existence, that is, the validity, of legal norms are subjected to these principles. Two statements of which one affirms the validity of a norm prescribing that men ought to behave in a certain way, and the other the validity of a norm prescribing that men ought not to behave in this way, contradict each other, just as two statements of which the one affirms that something is, and the other that it is not. If the one is true the other must be false. Two conflicting norms can be described as valid norms only by statements which contradict each other. In this sense we may say of conflicting norms that they "contradict" each other. Consequently two conflicting norms cannot be considered to be valid at the same time. Thus the science of law conceives of its object as a logical unit: a system of noncontradictory norms.

Although law is a norm, not a fact, there is nevertheless an essential relationship between norm and fact. The norm is — as pointed out — the meaning of a fact, the fact by which the norm is established. The fact by which a norm is established or, metaphorically speaking, created, is the act of an individual or a series of such acts intentionally directed at the behavior of another. If it is a legislative act or custom, it is a general norm. If it is a judicial or administrative act, it is an individual norm. The norm-creating act is a fact which exists in time and space and can be perceived by our senses. This fact can be described in an *is*-statement. But this fact is different from its meaning — that is, the norm — which is the object of jurisprudence, and which cannot be described in an *is*-statement, but only in an *ought*-statement. It is true that jurisprudence refers also to the procedure by which the legal norms are created; but only insofar as this procedure is prescribed or authorized by legal norms. The law regulates its own creation. Only the norms that prescribe or authorize the norm-creating acts are the object of jurisprudence. Jurisprudence deals with the legislative process only insofar as that process is determined in the constitution, and with the judicial and administrative processes only insofar as they are determined in statutes or rules of customary law.

The constitution, the statutes, the customary law are norms, and only as such the object of jurisprudence.

That a legal norm, in order to exist — and that means: in order to be valid — must be created by an act, which is a fact existing in time and space, is not the only relationship between norm and fact. A norm may or may not be obeyed and applied by a certain human behavior which actually takes place in time and space. A legal order as a whole and the particular legal norms which form this legal order are to be considered valid only if they are, by and large, obeyed and applied, only if they are effective. But their validity must not be confused with their effectiveness. Effectiveness is merely a condition of, but not identical with, validity. A legal norm may be valid before it becomes effective. When a statute is applied by a court for the first time after its adoption by the legislative organ, hence before the statute could become effective, the court applies a valid law; it can apply the law only if the law is valid. But the statute loses its validity if it has not become, or when it ceases to be, effective. The fact that a legal norm becomes effective must be added to the fact that it is created by an act; otherwise it can no longer be considered as valid. But just as the act by which the norm is created is not identical with the norm — which is the meaning of this act — the effectiveness of a legal norm is not identical with its validity.

The doctrine which defines law as a fact is based on the erroneous identification of the norm with the act whose meaning the norm is, and of the validity of the norm with its effectiveness. By avoiding this erroneous identification, the Pure Theory of Law separates jurisprudence, describing norms in *ought*-statements, from natural science describing facts in *is*-statements. This is the first reason that it is called a "pure" theory of law. The second reason is that it separates jurisprudence from ethics. The science of ethics describes norms, as does jurisprudence, but the norms described by ethics are not legal norms, but rather moral norms. The difference between legal and moral norms consists in that the former prescribe a certain behavior by attaching to the contrary behavior a coercive act as a sanction. A sanction is a forcible deprivation of life, freedom, property, or other values, as a reaction against a behavior considered by the legal authority as harmful to society. Law forbids murder, theft and the like by prescribing that if someone commits murder or theft, he ought to be punished by capital punishment or imprisonment. Law commands payment of one's debts by prescribing that if a person does not pay his debts, civil execution ought to be directed against his property. By attaching a sanction to a certain behavior, law qualifies this behavior

as a delict, as illegal, and makes the contrary behavior the content of a legal obligation. In this sense, law is a coercive order. Moral norms too, forbid or command a certain behavior, and some of them prescribe the same behavior as law, but without attaching to the contrary behavior a coercive act as a sanction. A moral order is a normative, but not a coercive order.

Insofar as moral norms are created by acts of human beings, by the founder of a religion, such as Moses, Jesus or Mohammed, or by custom, there are many different moral orders, valid at different times within different societies. Then the moral value has only a relative character. A definite legal order may or may not correspond to a certain moral order; but the validity of the legal order does not depend on its correspondence to a moral order. It may, from the point of view of a certain moral order, be considered as morally good or morally bad. That a legal order is morally good, means that it is just; that a legal order is morally bad, means that it is unjust. As a matter of fact, each legal order corresponds more or less to a definite moral order prevailing within a ruling group whose interests determine the law-creating process; and hence is considered, from the point of view of this moral order, as just. But at the same time it may be more or less contrary to a moral order prevailing within other groups, even of the same society for which the legal order is valid. Hence law must be distinguished from justice. A statement about the law must not imply any judgment about the moral value of the law, about its justice or injustice; which, of course, does not exclude the postulate that the law should be just. However, since there are not one justice, but many different and even contradictory ideals of justice, the postulate should never be raised without specifying which of these many justices is meant. Exactly as the Pure Theory of Law separates law from nature and thus jurisprudence from natural science, it separates law from morals, and thus jurisprudence from ethics. In these two respects it is a "pure" theory of law.

The Pure Theory of Law is a theory of positive law. Positive law is a coercive order whose norms are created by acts of human beings — by legislative, judicial or administrative agencies, or by custom constituted by acts of human beings. Legislative, judicial, and administrative agencies are human beings in their capacity as organs of the state. In so far as the law is created by these agencies, we say that the law is created by the state. We interpret the law-creating acts of definite human beings as "acts of state"; we attribute the function actually performed by a definite human being to the state. But, what is the state, and what is its relation to the law? The Pure Theory of Law shows that an act performed by a human being is interpreted as act of state only if this act is

determined in a specific way by the legal order; that to attribute the performance of this law-creating act to the state means to refer the act to the legal order by which the act is determined; that the state as an acting person creating the law is nothing else but the personification of the legal order which regulates its own creation; and that a state as a social order or political organization is this coercive order we call law; that a state imagined as a real being different from the law is the hypostatization of this relatively centralized legal order, or of its personification. Thus the Pure Theory of Law dissolves the misleading dualism of state and law prevailing in the traditional legal and political theory.

The Pure Theory of Law takes into consideration only positive law, norms created by acts of human beings. It does not take into consideration norms emanating from other, *i.e.*, superhuman authorities. Therefore it excludes from the province of jurisprudence any divine law, *i.e.*, law supposed to be created by God or a godlike entity. Consequently it excludes also so-called natural law, law which — according to the natural-law doctrine — is immanent in nature. Law — meaning norms regulating human behavior — can be immanent in nature only if nature is considered to be a legislator. To consider nature as a legislator implies to attribute a will to nature. This is either an animistic superstitution or a theological interpretation of nature, based on the belief that nature is created by God and hence a manifestation of God's absolutely good will. This is a metaphysical assumption incompatible with any scientific cognition and hence also with a science of law.

The same objection applies to that version of the natural-law doctrine which pretends to find the just norms of human behavior in the nature of man, particularly in his reason. Reason is the faculty of cognition. By our reason we are able to know, to understand or comprehend something which is given as an object of cognition, independently of this mental operation. To set a norm, to prescribe something, is a function of will, and human will is a psychic phenomenon totally different from human reason. Human reason can know norms after they have been created by acts of human will, but it cannot create norms. Only in God may reason and will be considered to coincide; only of God can people believe that knowing what ought to be is identical with willing that it ought to be. It is the old myth of the tree of knowledge: when men eat of the fruit of this tree, they will be like God, knowing good and evil. To God, knowing good and evil is the same as commanding the good and forbidding the evil. Only if the reason of man, as a being created by God in His own image, is part of the divine reason, can it be considered to be at the same time a norm-creating will, a so-called "practical reason." This self-contradictory notion,

which is the basis of a law of reason, is of theological origin.

In opposition to a theological or natural law doctrine the Pure Theory of Law, as a positivistic theory, does not see the reason for the validity of positive law in a divine or natural order, different from and above the positive law. It rejects the view that a positive law is valid only if its content corresponds to a divine or natural order, *i.e.*, valid only if the positive law is just. It considers every positive law — every coercive order which is established by acts of human beings and is by and large effective — as valid, without regard to its justice or injustice. The question of the reason for the validity of positive law — the question why the norms of any coercive order ought to be obeyed and applied — is, according to the Pure Theory of Law, to be understood as the following question: What is the logical condition under which the subjective meaning of the law-creating acts — that men ought to behave in a certain way — can be interpreted as their objective meaning? In answering this question we must be aware that it is logically impossible to infer from the statement that something is or is done, the statement that something ought to be or to be done, just as it is logically impossible to infer from the statement that something ought to be or to be done, the statement that something is or is done. This logical principle applies also to the fact of an act of will whose subjective meaning is that something ought to be done. From the fact that an individual commands that another individual ought to behave in a certain way, it does not follow that the other individual ought to behave in this way. It does not follow from the subjective meaning of the act of command that the meaning of the act is an objectively valid norm, disobedience to which would constitute something wrong. If a gangster commands that another person pay him a certain sum of money, we do not assume that this individual ought to obey the command and that, if he does not obey, he commits something wrong. A norm cannot be deduced from a fact; it can be deduced only from a norm. Hence the reason for the validity of a judicial decision or an administrative command is not the fact that a judge has actually rendered the decision, or an administrative organ has actually issued the command, but the statute authorizing the judge to render decisions and the administrative organ to issue commands. The reason for the validity of statutes is the constitution, authorizing an individual or a body of individuals to issue statutes. If it is historically a first constitution, and if the reason for the validity of this constitution cannot — from the point of view of a positivistic theory of law — be considered to be a superior order created by a divine, superhuman will, authorizing a certain individual or a body of individuals to establish the constitution, the reason for the validity of the con-

stitution and hence of the statutes, judicial decisions, and administrative commands established on the basis of the constitution can only be a norm we *presuppose*, if we are to interpret the acts whose subjective meaning the constitution, the statutes, the judicial decisions, the administrative commands are, as objectively valid norms. A norm is presupposed according to which men ought to behave in conformity with the constitution, hence in conformity with the general norms issued on the basis of the constitution by legislation or custom and, finally, in conformity with the individual norms issued on the basis of statutes or customary law by judicial and administrative acts; that is to say, in conformity with the legal order in its *hierarchical structure*. This norm, which is not a positive norm — not a norm created by an act of human or superhuman will, but only *presupposed* in juristic thinking — is the reason for the validity of a positive legal order. It is called the *basic norm*. Its presupposition is the condition under which every coercive order established by acts of human beings and by and large effective, may be interpreted as a system of objectively valid norms.

This presupposition is possible but not necessary. If the basic norm is not presupposed, a coercive order established by acts of human beings and by and large effective cannot be interpreted as a system of valid norms, but only as an aggregate of commands; and the relations constituted by such an order cannot be interpreted as legal relations, that is, as obligations, rights, competences and the like, but only as power relations. Thus the Pure Theory of Law, by ascertaining the basic norm as the logical condition under which a coercive order may be interpreted as valid positive law, furnishes only a conditional, not a categorical, foundation of the validity of positive law. Since the basic norm refers to a definite coercive order established by definite acts of human beings and by and large effective at a definite time and within a definite space, there is no choice between different basic norms. Only a basic norm referring to such a coercive order can be presupposed.

According to the Pure Theory of Law the basic norm may be presupposed with reference to every coercive order established by acts of human beings and by and large effective, whatever its content may be, that is to say, without regard to its justice or injustice. This theory does not aim at a moral or political justification of positive law. As a science of positive law, it refuses on principle to evaluate its object as just or unjust; it is unable to furnish a firm, an absolute, standard for evaluation. Those who expect such standard from jurisprudence may consider a positivistic theory of law in general and the Pure Theory of Law in particular as unsatisfactory. But their expectation can be fulfilled only by theological speculation: by metaphysics, not by science of law.

KELSEN VISITED

H. L. A. Hart*

In November, 1961, I had the enjoyable and instructive experience of meeting Hans Kelsen and debating with him at the Law School of the University of California in Berkeley some topics which I had previously selected for discussion from his *General Theory of Law and State*.[1] The meeting was arranged by Professor Albert Ehrenzweig who introduced us. We warned our very large audience that they might be disappointed or bored or both disappointed and bored: for the questions we proposed to discuss might excusably appear to them to be dry and technical, and our differences to be mere disputes over detail within the "positivist" camp of jurisprudence, of no great interest to those outside it. I explained that my view was that Kelsen's great work deserved the compliment of detailed scrutiny, and that it had too often been used as an excuse for the debate of vast and vaguely defined issues, such as the hoary perennial known as "Natural Law versus Legal Positivism." In spite of the technical nature of our discussion it was I think enjoyed by our audience which included, as well as lawyers, a sprinkling of philosophers, political theorists and students of other disciplines. Certainly it proved most instructive to me: it made me understand better the point of certain Kelsenian doctrines which had long perplexed me, even if it did not finally dispel my perplexities. I am reluctant to believe that I am alone in finding these difficulties in Kelsen's work; so some account of our discussion may be of use to others. In what follows I shall try to explain both why the points I raised seem to me important as well as to delineate our respective positions.

The points which I chose for discussion were these:

I. Kelsen's expression: "Rules of law in a descriptive sense."[2]
II. The definition of delict.[3]
III. The relationship between Positive Law and Morality.[4]

Besides these three issues there were others which we agreed to

* Professor of Jurisprudence, University of Oxford.

[1] KELSEN, GENERAL THEORY OF LAW AND STATE (1949). This work is referred to in this article as the *General Theory*.
[2] *Id.* at 45-46, 50, 163-64.
[3] *Id.* at 54-56.
[4] *Id.* at 373-76, 407-10.

alternative I proffered; for, again like the interpreter's words, the statements of the jurist representing the law are a specific kind of use of language and not a mention of it.

II. The Definition of Delict

Kelsen offers in his book what he terms a "juristic definition" of delict or, as English and American lawyers would say, of civil and criminal wrongs. In our debate I discussed this definition only so far as it related to crime and I was mainly concerned with the following quotations from the *General Theory*. These seem to me important because they show that Kelsen's pure theory differs from the usual conception of analytical jurisprudence in certain further respects beyond those already discussed above. They also seem to me to suggest certain limitations on the capacity of the Pure Theory to further the aim which Kelsen attributes to it of promoting the understanding of a system of positive law.

> From a purely juristic point of view, the delict is characterized as a condition of the sanction. But the delict is not the only condition. . . . What then is the distinctive characteristic of that condition which is called the "delict"? Could no other criterion be found than the supposed fact that the legislator desires conduct contrary to that which is characterized as "delict," then the concept of delict would be incapable of a juristic definition. The concept of delict defined simply as socially undesired behavior is a moral or political, in short, no juristic but a metajuristic, concept. . . .[17]
>
> A juristic definition of delict must be based entirely upon the legal norm. And such a definition can in fact be given. Normally, the delict is the behavior of that individual against whom the sanction as a consequence of his behavior is directed. . . . The criterion of the concept of "delict" is an element which constitutes the content of the legal norm. . . . It is an element of the norm by which the legislator expresses his intention in an objectively cognizable way; it is an element which can be found by an analysis of the legal norm. . . .
>
> The definition of delict as the behavior of the individual against whom the sanction, as a consequence of his behavior, is directed presupposes—although it does not refer to the fact—that the sanction is directed against the individual whose behavior the legislator considers to be detrimental to society . . . .[18]
>
> . . . . The legal concept of delict presupposes in principle that the individual whose behavior has from a political point of view a socially detrimental character, and the individual against whom the sanction is directly or indirectly executed, coincide. Only on this condition is the juristic definition of the delict, as the behavior of the individual against whom the sanction as a consequence of this behavior is directed, correct.[19]

[17] Kelsen, *op. cit. supra* note 1, at 53.
[18] *Id.* at 54.
[19] *Id.* at 56.

The general outline of this definition of delict is clear: a delict, *e.g.*, a crime, is simply the behavior upon which according to law a sanction becomes applicable to the person whose behavior it is. What is not clear is what Kelsen means by insisting on the one hand that this is all that a juristic definition of delict can and should say and, on the other hand, by acknowledging that this definition presupposes, though it does not refer to the socially detrimental character of the delict and is only correct if the condition thus presupposed is satisfied.

It is of course plain from many passages in Kelsen's book (and it is an important fact) that the Pure Theory imposes certain very severe restrictive conditions on the permissible forms of definition. It seems also clear that a science of positive law which disregarded these would not for Kelsen be a "normative" science. These restrictions indeed constitute one reason why no simple identification between analytical jurisprudence and either the Pure Theory or a "normative science" of law can be made in spite of their similarity in spirit and general orientation. For though Austin and his followers distinguish as sharply as Kelsen does between the analysis of law and moral, political or ideological evaluations of it, there is no counterpart in their work to Kelsen's distinctive insistence that in defining or analysing only certain restricted elements may be used. In general the Pure Theory insists that the clarificatory task of a normative science of law be performed with elements drawn from the law itself and care must be taken in defining or analyzing legal concepts to avoid using moral, political or psychological elements which are not, in Kelsen's words, "part of the legal material."

It is not very easy to make out precisely what elements these restrictions allow, but there are clear examples in Kelsen's book of what they exclude. Thus in criticizing Austin's analysis or definition of legal obligation Kelsen considers the definition that to be obliged is to fear the sanction, but he does not simply treat this, as a modern analytical jurist might, as an example of a mistaken definition. So he does not, for example, criticize it on the footing that a person may very well be under a legal obligation and yet not fear a sanction. What he does say is that such definition is "incompatible with the principles of analytical jurisprudence"[20] because "no analysis of the contents of commands can establish the psychological fact of fear."[21] His point is that it is wrong in principle to bring into the juristic definition of a concept psychological elements such as fear, or other elements which are not part of the contents of the law.

[20] *Id.* at 72.
[21] *Id.* at 72-73.

Kelsen's own juristic definition of obligation states that legal duty is "the behavior by the observation of which the delict is avoided, thus the opposite of the behavior which forms the condition of the sanction." No doubt Kelsen thinks that this definition is correct as complying with the restrictive condition that a juristic definition may use only elements which form part of the content of the law. It is worth noting, however, in order to prevent a common misunderstanding that though Kelsen rejects Austin's "psychological" conception of duty or obligation, he does not mean that a juristic definition can never use any psychological element. For Kelsen expressly says that in a case where the law itself makes such elements relevant, *e.g.*, where *mens rea* is a condition of criminal responsibility, then the sanction is directed to a psychologically qualified delict. The idea of responsibility based on fault is thus defined by Kelsen and no doubt he would claim that it is a sound juristic definition because though it uses psychological terms these are elements found in the relevant law.[22]

Though these examples throw some light on Kelsen's restricted form of juristic definition, it is not easy to understand why, given the aims of the Pure Theory, the restrictions it imposes should be observed; nor precisely how we are to determine what elements are to count as "found by an analysis of the content of the legal norm"[23] or "are expressed in the content of the norm"[24] or "are expressed in the material produced in the law-creating procedure"[25] or are "manifested in the contents of the legal order."[26] Kelsen certainly does insist that we must not bring into the definition of delict such elements as the supposed desire of the legislator, or the fact that the delictual conduct is socially harmful or against the purpose of the law: the juristic definition of delict must be "based entirely upon the legal norm"[27] and he considers his own definition of it to be so based. But this leaves much unexplained. Suppose that in fact the laws of a given system always contained (as Bentham wished) an explanatory statement that the actions to which the law attached criminal sanctions were regarded as social evils and that was why they were punished. Would the juristic definition of delict then rightly include a reference to such social facts? I am fairly sure that Kelsen's answer would be "No" though I much regret not having raised this point with him. He would, I think, in consistency with his general doctrine have to say that the laws of an actual system,

[22] *Id.* at 55, 66.
[23] *Id.* at 54.
[24] *Ibid.*
[25] *Id.* at 51.
[26] *Ibid.*
[27] *Id.* at 54.

before they have passed through the clarifying filter of the normative science of law, contain much that is irrelevant to that science. For the representation or description of the law which is the concern of that science is concerned only with its strictly normative elements; that indeed is why it is, in spite of Professor Alf Ross' protests, properly called a "normative science" and not merely a science *of* norms. I think this means that the permitted elements which may be used in juristic definition are those contained in the canonical form for the representation of the law which Kelsen lays down: statements that if such and such conditions are fulfilled then such and such a sanction shall follow. These are the statements by which the normative science of law is said by Kelsen to describe or represent law. They are "hypothetical judgment[s] attaching certain consequences to certain conditions":[28] if *A* is, *B* ought to be. So the explanatory statement of the law's purpose which would have pleased Bentham would, even if it were contained in the text of a statute, be quite irrelevant to normative science.

At this point Kelsen's restrictive conception of juristic definition can be seen to have points of contact with some themes of American legal realism. We might compare the restrictions insisted on by Kelsen to Holmes' "bad man" theory[29] that we should include in our definition, *e.g.*, of duty, only those elements which the "bad man" would want to know. Of course the permitted elements are quite different in the case of the two theories. The realist permits only elements relevant to the *prediction* of the sanction; whereas Kelsen permits only elements which according to the legal rule are conditions under which the sanction "ought" to be applied. But notwithstanding these differences the comparison does suggest a criticism of Kelsen's definition of delict and indeed of the whole program of his severely restricted juristic definition.

Briefly the criticism is that such definitions will not serve any useful purpose theoretical or practical and may introduce at points a confusion. That confusion may be generated is perhaps evident from the following simple case. Sanctions may take the form of compulsory money payments, *e.g.*, fines; but taxes also take this form. In both cases alike to use Kelsen's terminology certain behavior of the subject is a condition under which an official or organ of the system ought to demand a money payment from the subject. So if we confine our attention to the contents of the law as represented in the canonical form "If *A* then *B* ought to be" it is impossible to distinguish a criminal law punishing behavior with a fine from a revenue law taxing certain activities. Both when the individual is

28 *Id.* at 45.
29 HOLMES, *The Path of the Law*, in COLLECTED LEGAL PAPERS 171 (1920).

taxed and when he is fined the law's provisions when cast into the Kelsenian canonical form are identical. Both cases are therefore cases of delict unless we distinguish between them by reference to something that escapes the net of the canonical form, *i.e.*, that the fine is a punishment for an activity officially condemned and the tax is not. It may perhaps be objected that a tax, though it consists of a compulsory money payment as some sanctions also do, is not a "sanction" and that Kelsen's juristic definition of delict refers to a "sanction." But this does not really avoid the difficulty but only defers it; for we shall have to step outside the limits of juristic definition in order to determine when a compulsory money payment is a sanction and when it is not. Presumably it is a sanction when it is intended as or assumed to be a punishment to discourage "socially undesired behavior"[30] to which it is attached; but this is precisely the element which Kelsen considers to be excluded from the juristic definition of delict.

It is plain that Kelsen himself is aware of these difficulties because he concedes that the juristic definition only holds good on the presupposition that the behavior, which is the condition of the sanction, is considered detrimental to society. But does not this concession show that the severely restricted juristic definition is useless as well as confusing? Here it is important to stress that many of the illuminating definitions of the Pure Theory are not and could not be *juristic* definitions in the severely restricted sense that Kelsen intends. Plainly for the reasons given above the definition of a sanction is not.[31] It is even possible to doubt whether the definition of a legal norm (quite apart from its dependence on the definition of a sanction) conforms to the strict requirements of juristic definition. For Kelsen tells us that the norm "is the expression of the idea that something ought to occur, especially that an individual ought to behave in a certain way."[32] But though a norm may *be* an expression of an idea it is not clear that "an expression" or "an idea" or "an expression of an idea" are *contents* or *elements* of the norm or fit any other of the descriptions given by Kelsen of what may be used in a strictly juristic definition. So we should perhaps distinguish the most fundamental definitions of the Pure Theory to which the jurist conducting the normative science of law will *conform* in representing the law of a particular system as "metajuristic" definitions, to mark the distinction between them

---

[30] KELSEN, *op. cit. supra* note 1, at 53. The difficulty of distinguishing a penalty from a tax for the purpose of art. I, § 8, of the United States Constitution is well known. See, *e.g.*, Steward Mach. Co. v. Davis, 301 U.S. 548 (1937).

[31] See the discussion of coercion and the distinction between civil and criminal sanctions in KELSEN, *op. cit. supra* note 1, at 18-19, 50-51.

[32] KELSEN, *op. cit. supra* note 1, at 36.

and the juristic definitions which the jurist will actually *use* in representing the law of some particular system. He will not use in his representation of the system, but will take for granted, definitions of "sanction" or of "legal rule" but he will use definitions of delict. Perhaps indeed some such distinction between definitions which are metajuristic and those which are juristic is needed for any analytical account of law.

I pressed these points on Kelsen in our debate but I cannot say that he retreated or was moved by my claim that he had in fact given his case away by saying that his definition of delict held good on the "presupposition in principle that the behavior against which the sanction is directed has or is considered to have a socially detrimental character." I did however learn from our discussion two important things. The first is that Kelsen had an interesting and possibly a good reason for talking not merely of a science of norms but of a "normative" science of law and this is not open to Professor Ross' criticism though it may be to others. The second is that any one who, like myself, would wish to bring into the definition of crime or delict the idea that the behavior to which sanctions are attached is unlike behavior which is simply taxed and differs from it because it is in some way condemned, must be careful to state how in the case of any given law the presence of this factor of condemnation is ascertained.

III. The Relationship Between Law and Morality

> Let us consider the case of a conflict between a norm of positive law and a norm of morality. Positive law can, for instance, stipulate an obligation to render military service, which implies the duty to kill in war, while morality, or a certain moral order, unconditionally forbids killing. Under such circumstances, the jurist would say that "morally, it may be forbidden to kill, but that is irrelevant legally." From the point of view of positive law as a system of valid norms, morality does not exist as such; or, in other words, morality does not count at all as a system of valid norms if positive law is considered as such a system. From this point of view, there exists a duty to perform military service, no contrary duty. In the same way, the moralist would say that "legally, one may be under the obligation to render military service and kill in war, but that is morally irrelevant." That is to say, law does not appear at all as a system of valid norms if we base our normative considerations on morality. From this point of view, there exists a duty to refuse military service, no contrary duty. Neither the jurist nor the moralist asserts that both normative systems are valid. The jurist ignores morality as a system of valid norms, just as the moralist ignores positive law as such a system. Neither from the one nor from the other point of view do there exist two duties simultaneously which contradict one another. And there is no third point of view.[38]

[38] *Id.* at 374.

Traditional Naturalism and Neo-Naturalism

Rebecca Prebble, 2003

I INTRODUCTION

A What is naturalism?

Naturalism is a school of thought that has persisted in different guises throughout the history of jurisprudence. Despite this tenacity, "naturalism" is a term that continues to frustrate attempts at comprehensive definition. In a loose sense, naturalism is the idea that there is some intrinsic link between law and morality, and that a proper legal system is necessarily consistent with some moral framework. This framework may be grounded in the presumed existence of a (probably Christian) God, or it may be based on elusive ideas of justice and fairness. The key idea is that a separation of law and morality is neither desirable nor possible.

B Analytical framework

Naturalism is not a theory that is easily divisible into smaller groupings of ideas. However, a distinction is conventionally made between naturalist thought before the twentieth century (traditional naturalism), and the neo-naturalist developments of the last hundred years. To highlight the differences and similarities between traditional naturalism and neo-naturalism, this essay first gives a brief overview of the key developments in traditional naturalist theories, and then examines how elements of these theories have been discarded, modified, or absorbed by neo-naturalism.

II HISTORICAL OVERVIEW

A Foundations

Naturalist thought can be traced back to Aristotle in the third century BCE. In his *Nicomachean Ethics*, Aristotle stipulates that justice is an essential requirement for any good society. He divides justice into two varieties: natural justice and conventional justice. Rules of conventional justice are similar to positive law in that they exist only because some specific authority has laid them down. In contrast, rules of natural justice necessarily exist because of

human rationality. Therefore they are recognised as universal rules by all rational entities.[1]

Cicero borrowed and expanded on Aristotle's division of law in the first century BCE. Cicero proposed three layers of law: divine, natural, and positive. Divine law was supreme, but not wholly accessible to human understanding. The aspects of it that humans could understand formed the natural laws. Positive law was the law laid down by the governing authority. Positive law was good law to the extent that it complied with natural law. It is in the philosophy of Cicero that we first come across the idea that citizens have the right to disobey a law that does not conform to some higher morality. Cicero even goes so far as to say that the concepts of justice and truth are inherent in the definition of law; so a command that does not conform to natural law is not law.[2]

B Medieval naturalism

Naturalism developed as a truly independent doctrine during the Middle Ages. Medieval jurists took the separation of natural and positive law and reformulated it to fit in with Christian thought. Thomas Aquinas founded his idea of a higher standard of morality in the laws of the Christian God, but he also held that these natural laws were accessible by the exercise of human reason. Aquinas's theory proposes a yardstick for ascertaining whether human law is valid: its validity depends on its justice, which in turn depends on whether it is in accordance with natural law.[3] Positive law is thus defined as "a rational ordering of things which concern the common good; promulgated by whoever is charged with the care of the community".[4] Aquinas reintroduces the definitional question of whether a bad law (one that is not in accordance

---

[1] Sir Frederick Pollock, "The History of the Law of Nature" in Sir Frederick Pollock (ed) *Jurisprudence and Legal Essays* (Macmillan & Co Ltd, London, 1961), 125.

[2] Cicero, "De Legibus", in JC Smith and David N Weisstub (eds) *The Western Idea of Law* (Butterworth & Co Ltd, Toronto, 1983), 343, 344.

[3] Thomas Aquinas, "Summa Theologica" in Lord Lloyd of Hampstead and MDA Freeman *Lloyd's Introduction to Jurisprudence* (5 ed, Stevens and Sons, London, 1985), 151, 155.

[4] Id, 152.

with natural law) is nevertheless still a law. Aquinas seems to hold that such a law is "a corruption of law,"[5] but he does not deny that in the strict sense such a corruption is still a law. A related question is whether there is any moral obligation to obey a bad law. Cicero is firm that there is no such obligation,[6] but Aquinas does not seem to reach a conclusion on this issue. There is clearly an obligation to follow natural law, and to follow positive law to the extent that it is in line with natural law. However, Aquinas does not definitively determine the status of our obligations regarding bad law.

C The early modern era

The decline in the power of the Catholic church had a great impact on jurisprudence. England's switch to Protestantism was a particularly dramatic signal of Catholicism's waning influence, but Continental philosophers also began to move away from a Catholic, or even Christian, foundation for legal theory. Philosophers' conception of natural law begins subtly to change in this period, reflecting contemporary dissatisfaction with ineffective government. While natural law had previously been seen as a tool that could be used to evaluate positive human law, it now emerges as a guide for effective society. The idea of the social contract was not totally new, but it is during this period that it emerges in a cohesive form.

John Locke was writing in the late seventeenth century, and his philosophy was necessarily influenced by contemporary political upheaval. Locke was concerned with putting natural law to practical use to a far greater extent than the theorists that preceded him. While previous theories had been fairly general in application, Locke took the idea further and tried to find naturalist justifications for the protection of property rights and the good government of society. Locke's conception of natural law presupposes the existence of a Christian god, but the influence of Christianity is not felt so forcefully in his

---

[5] Id, 155.

[6] Cicero, "De Legibus", in JC Smith and David N Weisstub (eds) *The Western Idea of Law* (Butterworth & Co Ltd, Toronto, 1983) 343, 344.

theory as it is in the theories of earlier jurists such as Aquinas. Locke's theory uses human reason as its key tenet: natural law follows not simply from God's existence, but from human *knowledge* of God's existence.[7] True to his empiricist roots, Locke holds that knowledge of natural law is not innate to humans, but must (and can) be learnt through experience and reason. While God's role in naturalism diminishes with Locke, it is still present. Our obligation to obey natural law comes from God's existence: "He has the power to enforce [natural law] through rewards and punishments."[8]

Locke formulated a naturalist justification for humans having property rights. His conceptions of property rights included such rights as the rights to life and liberty. For Locke, property is the result of labour, and labour is in accordance with the commands of God and natural law; so natural law should protect property rights.[9] It is in this idea that we see the beginnings of the social contract, a theme that runs through much of naturalist theory in one form or another. Locke holds that people choose to enter into a society in order that their rights will be protected, and in doing so allow the sovereign of that society to have power over them.[10] Locke's theory is similar to many pre-nineteenth century naturalist theories in that it falls somewhere between being normative and being descriptive. His theory of society appears to be an aspirational ideal, rather than a description of what actually occurs. However, he also holds that a sovereign's commands that are not in accordance with natural laws are neither good nor valid,[11] suggesting that such commands are not law at all.

Thomas Hobbes, writing slightly before Locke, saw the social contract structure as a means of controlling people, as well as a means of protecting

---

[7] John Locke, *An Essay Concerning Human Understanding* (1690) Ed P H Nidditch (Clarendon Press, Oxford, 1975), I.3.6.

[8] Id, 8.

[9] John Locke, Two Treatises of Government, in Dennis Lloyd *Introduction to Jurisprudence* (2 ed, Stevens & Sons, London, 1965) 80.

[10] Id, 81.

[11] Id, 82.

people from foreign threats. The essence of Hobbes's social contract was that the will of the people was to be implemented by the sovereign.[12] Like Locke, Hobbes based his social contract on natural laws, but instead of Locke's rights-based approach, Hobbes proposed two fundamental laws. First, people were to strive for peace, and only use violence in self defence. Secondly, people should allow others the freedom that they would want to have themselves.[13] It is arguable that Hobbes's theory relies even less on some form of divine law than Locke's. Hobbes proposes the social contract as a way to ensure humans can live together in freedom: there is little suggestion that in so doing they are acting in accordance with God's will.

Jean-Jacques Rousseau, writing immediately prior to the French Revolution in 1779, developed the idea of the social contract further beyond its Christian roots. The purpose of such an arrangement is the preservation of individual liberties, but these liberties are portrayed as innate, deriving from an original "state of nature", rather than any divine power.[14] However, it is clear that God still has a role in such a society, because for Rousseau God is the source of justice.[15] Rousseau held that individuals entering into the social contract did so with other individuals. The contract was not with the sovereign: the role of the sovereign was simply to give effect to the will of the populace. If the sovereign does not perform its role adequately, the populace may overthrow it, that is, they may collectively opt out of the social contract.[16] Rousseau is typical of the jurists of his period in that he uses naturalism not so much to assess individual laws but to found a theory of the good government of society. Naturalism can therefore be seen as a more normative theory than it had been up to this point. Rousseau and his contemporaries were proposing an aspirational society: the context of their work makes it clear that they did not

---

[12] Thomas Hobbes, "Leviathan", in Lord Lloyd of Hampstead and MDA Freeman (eds) *Lloyd's Introduction to Jurisprudence* (5 ed, Stevens & Sons, London, 1985), 156, 157.

[13] Id, 157-158.

[14] Jean-Jacques Rousseau, *The Social Contract*, (translated by Maurice Cranston) (Hazell Watson & Viney Ltd, London, 1968) 64.

[15] Id, 80.

[16] Id, 145.

consider themselves to be living in such a society. However, it is unlikely that they would deny that they lived in a "legal" system. The descriptive element of naturalist theory is therefore temporarily lost.

D Neo-naturalism

The nineteenth century saw naturalism all but disappear, and the positivism of Bentham and Austin became the prevailing theory. In retrospect, the popularity of positivism at that time was merely a pause in the history of naturalism, rather than a sign of the theory's demise. The emergence of positivism is closely linked to the decline of Christianity: people realised that religion was inadequate to explain law; so the realism of positivism was attractive. The revival of naturalism in the twentieth century is partly a response to the emergence of totalitarian regimes at the beginning of the century. Nazi and Bolshevik law satisfied the definition of law by the tenets of positivism, but people found these legal systems appalling. A theory of law without a moral character suddenly seemed inadequate, and the neo-naturalists emerged.

Theorists like Lon Fuller and Ronald Dworkin, and to some extent John Rawls, take traditional naturalist ideas and modify them to suit current ideas. Fuller bases his naturalism on the procedural requirements that law must have, while Dworkin takes a rights-based approach, and stresses the importance if the preservation of natural rights. Neo-naturalism has strong ties to traditional naturalism, but many of the distinguishing features of naturalist thought are lost or substantially modified in the twentieth century. In particular, religious belief as a moral base is lost, as is the traditional naturalist view of seeing society as a contract with attendant mutual rights and obligations. In reaction to positivist criticism, we also see naturalism definitely becoming a descriptive theory, as opposed to an arguably normative one.

III TRADITIONAL NATURALISM AND NEO-NATURALISM

A Moral framework

Any theory that purports to be a naturalist one must operate within some sort of moral framework. That is, a naturalist must provide some objective standard against which to evaluate laws and judicial decisions. A key difference between traditional naturalism and neo-naturalism is that neo-naturalism does not rely on any sort of divine law for its foundation. Nevertheless, naturalism requires some sort of morality with which to replace religion. Thus a main focus of neo-naturalist theory has been the establishment and justification of an alternative moral framework.

Lon Fuller is an example of a jurist who proposes a moral framework that has little to do with Christian morality. In fact, Fuller proposes a system that mentions substantive morality only obliquely. Fuller focuses on what he calls law's "internal morality."[17] He holds that if the commands of a sovereign are to be called law at all (aside from any question over whether they are "good"), they must have eight properties. They must be general, intelligible, prospective, consistent, congruent, published, practicable and constant. Fuller goes to great pains to point out that while his theory is one of natural law, it is not based on any idea of divinity, or any sort of higher moral standard.[18] The rules that he proposes are certainly fundamental rules to which law must conform, but he does not specify any substantive morality that serves the same purpose. In contrast with Locke, for example, who specifies that one of the things that law must do is preserve individual property rights, Fuller makes no claims about the substantive aims of law; so Fuller's moral framework departs completely from the traditional naturalist conception: not only does he propose different objective standards, he proposes a different *variety* of standard, that is, formal as opposed to substantive.

[17] Lon Fuller, *The Morality of Law* (Yale University Press, New Haven (CT), 1964) 171.

[18] Id, 170.

Ronald Dworkin takes a different approach. Like Fuller's, Dworkin's moral framework could be classed as "internal", but Dworkin's morality has a more substantive aspect. Dworkin proposes that objective principles exist in the fabric of the law.[19] These principles are not necessarily based in an identifiable morality, although some of them might be. Dworkin does not attempt to list all legal principles, and in fact does not believe such an enterprise is possible, but gives us the example of "a man may not profit from his own wrong".[20] Dworkin's principles form a moral framework because they provide standards by which we may assess law, particularly judicial decisions. Principles are not external standards, they are as much a part of the law as legislatively enacted rules.[21] Dworkin also proposes that natural rights form a sort of substantive morality. We can assess the merit of a law or other act of government with reference to whether it infringes individual rights.[22] Dworkin's theory perhaps stays closer to the traditional type of moral framework than Fuller's: Dworkin proposes that there is some sort of right and wrong within the law, a right and wrong that we can assess with reference to substantive values. However, these substantive values have no link with Christianity. Rights against the government are not God-given, yet they are still fundamental.[23]

The lack of a Christianity-based moral framework is the most obvious factor that distinguishes neo-naturalism from traditional naturalism, but it would be incorrect to conclude that the acceptance of a secular framework was a sudden change that arrived with the twentieth century naturalists. Rather, an examination of the work of Locke and Rousseau reveals that Christianity's influence was fading in even the early modern era. While earlier theorists like Aquinas rooted their theories firmly in Christianity, Locke and Rousseau place

[19] Ronald Dworkin, *Taking Rights Seriously* (Harvard University Press, Boston, 1978) 24.

[20] Id, 27.

[21] Id, 40.

[22] Ronald Dworkin, "Taking Rights Seriously" in JC Smith and David N Weisstub (Eds) *The Western Idea of Law* (Butterworth & Co Ltd, Toronto, 1983) 490, 497.

[23] Id, 497.

more reliance on human reason and the obligations individuals have towards each other. Neo-naturalism's discarding of a Christian framework is therefore the culmination of a long process, as opposed to a brand new idea.

The fundamental problem of a Christian moral framework, and the motivation for its eventual rejection, is that a Christian framework gives non-Christians no reason to accept naturalism as a valid theory. Ideally, the norms natural law proposes should be seen as desirable regardless of any particular ontology. Modern theorists recognise that the moral framework of their theories must be independent of any appeal to religious values; so neo-naturalism typically disclaims any connection with Christianity.[24]

B Normative/descriptive

Traditional naturalists did not make it clear whether their theories were normative or descriptive. Close analysis reveals them to tend towards being normative, particularly in the case of the theorists of the early modern era. Locke, Rousseau and their contemporaries focused on how a society should be governed. They set standards for how a law could be evaluated and how a good society should operate. Their purpose was to point out the deficiencies in the legal systems of the time, as opposed to formulating a definition of law itself. Positivists such as HLA Hart have attacked this normative approach by arguing that naturalism evades the question of what the law is by instead saying what it ought to be.[25] The positive conception of what makes a command a law leaves no room for a moral element. To refute positivist criticism, neo-naturalists must be clear that their theories are not simply normative, and that they make real claims about the nature of law as it is.

This descriptive element comes through in the work of Lon Fuller. It is true that his eight properties of law all have a moral element in that we would

---

[24] Not all twentieth century naturalists reject Christianity as a moral framework. For example, John Finnis argues that Christian morality has a place in a legal framework. See John Finnis, *Natural Law and Natural Rights* (Oxford University Press, New York, 1980).

[25] HLA Hart, "The Separation of Law and Morals" (1958) 71 Harv L Rev 593, 599.

consider it desirable that a law should have them. Nevertheless, Fuller holds that his theory is not merely descriptive or aspirational: his eight properties are properties of everything that can rightly be called law, not just good law.[26]

Ronald Dworkin's work reveals a similar determination to be descriptive. Dworkin holds that his principles are as much a part of the law as enacted rules are. According to Dworkin, if a judge makes a decision without taking principles into account, that judge may be criticised not only because the decision may seem morally wrong, but because the decision was not made according to law.[27]

C Societal structure

Assessment and definition of individual laws is just one of the two main issues that naturalism has confronted. Naturalist theory also looks at legal systems generally, and provides guidelines for the efficient running of society. The Locke-Hobbes-Rousseau idea of the social contract has had a profound influence on the way we think about society. The concept has two key elements: first, the description of a legal system as a kind of contract, and secondly, the terms of that contract and the parties' rights under it. Neo-naturalists tend to modify at least one element when prescribing requirements for a good society, but the basic framework is still discernible.

John Rawls sees the social contract as mainly involving obligations on the side of the government, although it is unlikely that he would deny that citizens have duties as well as rights. While the traditional purpose of a social contract is for the preservation of rights, Rawls's purpose emerges as the implementation of the principles of justice.[28] Rawls proposes two main such principles: first, justice is equality in the assignment of basic rights and duties,

---

[26] Lon Fuller, *The Morality of Law* (Yale University Press, New Haven (CT), 1964) 169.

[27] Ronald Dworkin, *Taking Rights Seriously* (Harvard University Press, Boston, 1978) 24.

[28] John Rawls "A Theory of Justice" in JC Smith and David N Weisstub (Eds) *The Western Idea of Law* (Toronto, Butterworth & Co Ltd, 1983), 479, 479.

and secondly, social and economic inequalities are justified only if they result in compensatory benefits for the whole of society.[29] Rawls uses the social contract as a framework for a wealth-redistributive model of society; his theory is thus a normative one that first presumes a certain social ordering. While Rawls may not be thought of primarily as a neo-naturalist, his theory clearly has naturalist aspects. His principles of justice, for instance, may be thought of natural law complexes. Rawls departs from traditional social contract ideals in that he does not specifically address the rights of the populace in circumstances where the sovereign breaks the contract. He recognises that membership of society is not strictly voluntary, as was assumed by Locke and Rousseau,[30] suggesting that the citizens' right to revolt is something of an anachronism today.

Fuller leads us away from the traditional social contract idea: he has a more pragmatic view of society that starts with the basic truth that people are subject to governmental control. Since there is little choice as to membership of a society, Fuller stresses the importance of the government's adherence to the procedural requirements of good lawmaking.[31] The power that a government has over its citizens is thus constrained. Fuller does not give criteria for substantive justice; he maintains that a society that maintains his eight principles will necessarily be substantively just.[32] While it may be conceivable that a substantively unjust legal system could still adhere to Fuller's eight principles, Fuller maintains that in the past gross injustices have always been accompanied by breaches of his procedural requirements. The lack of contractual structure in Fuller's theory of governmental power means that he would be unlikely to sanction citizens disobeying a law that they found substantively unjust. Fuller puts a high value on fidelity to duly enacted law,

[29] Id, 481.

[30] Id, 480.

[31] Lon Fuller, *The Morality of Law* (Yale University Press, New Haven (CT), 1964) 169-170.

[32] Id, 170

and seems to regard minor infringements of his procedural requirements as a lesser evil than public disorder.[33]

Fuller's approach may be contrasted with Dworkin's. Dworkin's conception of a good society is one in which the government does not encroach on the individual rights of citizens.[34] Unlike other neo-naturalists, Dworkin allows citizens to have some remedies if the government does not recognise their rights. While he does not go so far as to say that the entire populace may revolt if the government behaves badly enough, Dworkin holds that it would not be wrong for individual citizens to disobey laws that infringe on their rights.

The social contract paradigm as originally formulated included in the early modern era included a right to revolt. The exclusion of this right in the theories of neo-naturalists therefore represents a significant change in outlook. The absence of the right to revolt in modern theories comes as a result of an increasing awareness of the consequences of allowing citizens such a right. Theorists like Rousseau and Locke could perhaps take much of the blame for some particularly bloody periods in European history. The French Revolution was a response to an unjust regime, but it is arguable that the cure was worse than the disease. The same could be said about the Russian Revolution. Later jurists are understandably more circumspect about allowing people such unbridled rights. Dworkin perhaps goes the furthest in allowing individuals to disobey specific unjust laws, but he draws the line at revolution.

The right to revolt is a logical consequence of viewing society as a contract, because with a contractual structure a breach by one side allows a breach by the other. It follows that modern naturalists have largely abandoned the contractual view of society in favour of a more practical and realistic

---

[33] Lon Fuller, "Positivism and Fidelity to Law – A Reply to Professor Hart" (1957) 71 Harv L Rev 632, 670.

[34] Ronald Dworkin, "Taking Rights Seriously" in JC Smith and David N Weisstub (Eds) *The Western Idea of Law* (Butterworth & Co Ltd. Toronto, 1983), 490 491.

approach. This more sophisticated perspective better accommodates the desirable idea that the government has duties towards its citizens irrespective of those citizens fulfilling any obligations.

The populace's right to revolt if they are dissatisfied with the performance of the sovereign is the aspect of the social contract idea that causes the most controversy. Theorists like Locke and Rousseau who assert that this right exists (as it logically must if the social contract is a correct representation of society) are vague about when it may be exercised. Rousseau gives some substantive criteria for the good exercise of government power, such as tranquility, and the preservation of individual freedoms and property rights,[35] but gives little indication of when minor infringements justify a revolution. Similarly, Locke sees freedom from arbitrary power and the protection of property rights as the marks of a legitimate political system,[36] but again is unhelpful in determining just when a government may be said to be acting in a manner that allows revolt.

[35] Jean-Jacques Rousseau, *The Social Contract*, (translated by Maurice Cranston) (Hazell Watson & Viney Ltd, London, 1968) 129.

[36] Stephen Buckle *Natural Law and the Theory of Property*, (Oxford University Press, Oxford 1991) 178.

A HISTORY OF POLITICAL THEORY[1]

George H. Sabine

Cicero and the Roman Lawyers

Cicero's true importance in the history of political thought lies in the fact that he gave to the Stoic doctrine of natural law a statement in which it was universally known throughout western Europe from his own day down to the nineteenth century. From him it passed to the Roman lawyers and not less to the fathers of the church. The most important passages were quoted times without number throughout the Middle Ages. It is a significant fact that, though the text of *The Republic* was lost after the twelfth century and not recovered until the nineteenth, its most striking passages had already been excerpted into the books of Augustine and Lactantius, and so had become matters of common knowledge. The ideas were, of course, in no sense original with Cicero but his statement of them, largely in Latin expressions of his own devising to render the Stoic Greek, became incomparably the most important single literary means for spreading them through western Europe. A few of Cicero's great passages must be kept in mind by anyone who wishes to read political philosophy in the centuries that followed.

First of all, there is a universal law of nature arising equally from the fact of God's providential government of the world and from the rational and social nature of human beings which makes them akin to God. This is, as it were, the constitution of the world-state ; it is the same everywhere and is unchangeably binding upon all men and all nations. No legislation that contravenes it is entitled to the name of law, for no ruler and no people can make right wrong :

> There is in fact a true law — namely, right reason — which is in accordance with nature, applies to all men, and is unchangeable and eternal. By its commands this law summons men to the performance of their duties ; by its prohibitions it restrains them from doing wrong. Its commands and prohibitions always influence good men, but are without effect upon the bad. To invalidate this law by human legislation is never morally right, nor is it permissible ever to restrict its operation, and to annul it wholly is impossible. Neither the senate nor the people can absolve

us from our obligation to obey this law, and it requires no Sextus Aelius to expound and interpret it. It will not lay down one rule at Rome and another at Athens, nor will it be one rule today and another tomorrow. But there will be one law, eternal and unchangeable, binding at all times upon all peoples ; and there will be, as it were, one common master and ruler of men, namely God, who is the author of this law, its interpreter, and its sponsor. The man who will not obey it will abandon his better self, and in denying the true nature of a man, will thereby suffer the severest of penalties, though he has escaped all the other consequences which men call punishments.[2]

In the light of this eternal law all men, as Cicero insists in the most unequivocal terms, are equal. They are not equal in learning, and it is not expedient for the state to try to equalize their property, but in the possession of reason, in their underlying psychological make-up, and in their general attitude toward what they believe to be honourable or base all men are alike. Indeed Cicero goes so far as to suggest that it is nothing but error, bad habits, and false opinions that prevents men from being in fact equal. All men and all races of men possess the same capacity for experience and for the same kinds of experience, and all are equally capable of discriminating between right and wrong.

Out of all the material of the philosophers' discussions, surely there comes nothing more valuable than the full realization that we are born for justice, and that right is based, not upon man's opinions, but upon nature. This fact will immediately be plain if you once get a clear conception of man's fellowship and union with his fellow men. For no single thing is so like another, so exactly its counterpart, as all of us are to one another, Nay, if bad habits and false beliefs did not twist the weaker minds and turn them in whatever direction they are inclined, no one would be so like his own self as all men would be like all others.[3]

Professor A.J. Carlyle has said that "no change in political theory is so startling in its completeness"[4] as the change from Aristotle to a passage such as this. The process of reasoning is, in truth, the exact opposite of that which Aristotle had used. The relation of free citizenship for Aristotle can hold only between equals, but because men are not equal, he had inferred that citizenship must be restricted to a small and carefully selected group. Cicero on the contrary infers that, because all men are subject to one law and so are fellow citizens, they must be in some sense equal. For Cicero equality is a moral requirement rather than a fact ; in ethical terms it expresses much the same conviction that a Christian might express by saying that God is no respecter of persons. There is no implication of political democracy, though without some such moral conviction democracy would be hard to defend. What is asserted is that some measure of human dignity and respect is due to every man ; he is inside and not outside the great human brotherhood. Even if he were a slave he would not be, as Aristotle had said, a living tool, but more nearly as Chrysippus had said, a wage-earner hired for life. Or, as Kant rephrased the old ideal eighteen centuries later, a man must be treated as an end and not as a means. The astonishing fact is that Chrysippus and Cicero are closer to Kant than they are to Aristotle.

The political deduction which Cicero draws from this ethical axiom is that a state cannot exist permanently, or at least cannot exist in any but a crippled condition, unless it depends upon, and acknowledges, and gives effect to the consciousness of mutual obligations and the mutual recognition of rights that bind

its citizens together. The state is a moral community, a group of persons who in common possess the state and its law. For this reason he calls the state, in a fine phrase, the *res populi* or the *res publica,* "the affair of the people," which is practically equivalent in meaning to the older English use of the word "commonwealth." This is the ground for Cicero's argument, against the Epicureans and Skeptics, that justice is an intrinsic good. Unless the state is a community for ethical purposes and unless it is held together by moral ties, it is nothing, as Augustine said later, except "highway robbery on a large scale." A state may of course be tyrannous and rule its subjects by brute force — the moral law does not make immorality impossible — but in the measure that it does so, it loses the true character of a state.

> The commonwealth, then, is the people's affair ; and the people is not every group of men, associated in any manner, but is the coming together of a considerable number of men who are united by a common agreement about law and rights and by the desire to participate in mutual advantages.[5]

The state, then, is a corporate body, membership in which is the common possession of all its citizens ; it exists to supply its members with the advantages of mutual aid and just government. Three consequences follow. First, since the state and its law is the common property of the people, its authority arises from the collective power of the people. A people is a self-governing organization which has necessarily the powers required to preserve itself and continue its existence : *Salus populi suprema lex esto.* Second, political power when rightfully and lawfully exercised really is the corporate power of the people. The magistrate who exercises it does so by virtue of his office ; his warrant is the law and he is the creature of the law.

> For as the laws govern the magistrate, so the magistrate governs the people, and it can truly be said that the magistrate is a speaking law, and the law a silent magistrate.[6]

Third, the state itself and its law is always subject to the law of God, or the moral or natural law — that higher rule of right which transcends human choice and human institution. Force is an incident in the nature of the state and is justified only because it is required to give effect to the principles of justice and right.

These general principles of government — that authority proceeds from the people, should be exercised only by warrant of law, and is justified only on moral grounds — achieved practically universal acceptance within comparatively a short time after Cicero wrote and remained commonplaces of political philosophy for many centuries. There was substantially no difference of opinion about them on the part of anyone in the whole course of the Middle Ages ; they became a part of the common heritage of political ideas. There might, however, be considerable differences of opinion about the application of them, even among men who had not the remotest doubt about the principles themselves. Thus everyone agrees that a tyrant is despicable and his tyranny a bitter wrong against his people, but it is not obvious just what the people are entitled to do about it, or who is to act in their behalf in doing it, or how bad the abuse must be before measures are justified. In particular, the derivation of political authority from the people does not of itself imply any of the democratic consequences which in modern times have been deduced from the consent of the governed. It does not say who

speaks for the people, how he becomes entitled to speak, or exactly who "the people" are for whom he speaks — all questions of the utmost practical importance. The use of the ancient principle that political authority comes from the people to defend the modern forms of representative government was merely the adaptation of an old idea to a new situation.

The classical period in the development of Roman jurisprudence fell in the second and third centuries after Christ, and the writings of the great jurists of that age were excerpted and compiled into the Digest (or Pandects), which the Emperor Justinian caused to be published in 533. The political philosophy which is embedded in this body of legal writing is a repetition and elaboration of the theories found in Cicero.

Political theory forms an insignificant proportion of the whole work, the relevant passages being neither very numerous nor very extensive. The lawyers were jurists, not philosophers. For this reason it is often hard to tell just how seriously a philosophical idea is to be taken when it occurs ; one does not know whether the writer himself regarded it as a polite embellishment or whether it really influenced his legal judgment. Obviously it was never part of the lawyers' purpose to formulate a political philosophy or to inject a philosophy into the law. The philosophy of the Roman lawyers was not philosophy in a technical sense but certain general social and ethical conceptions, known to all intelligent men, which were in some way considered to be useful for their own juristic purposes. This makes it the more striking that they uniformly selected philosophical ideas belonging in the Stoic and Ciceronian tradition. The ideas of egoistic individualism, contained in the writings of the Epicureans and the Skeptics, must have been equally at their disposal, but the lawyers found no use for them. The fact that their interest in political theory was desultory and unsystematic does not mean that what they had to say was unimportant. The enormous authority attached to the Roman law throughout western Europe gave weight to any proposition which was a recognized part of it. Moreover, any general conception embedded in the law was certain to be known to all educated men as well as to lawyers, and ultimately by common report to many who were not scholars at all. In the end the Roman law became one of the greatest intellectual forces in the history of European civilization, because it provided principles and categories in terms of which men thought about all sorts of subjects and not least about politics. Legalist argumentation — reasoning in terms of men's rights and of the justifiable powers of rulers — became and remained a generally accepted method of political theorizing.

The lawyers excerpted in the Digest, as well as those who formulated Justinian's Institutes in the sixth century, recognized three main types of law, the *ius civile*, the *ius gentium* and the *ius naturale*. The *ius civile* is, of course, the enactments or the customary law of a particular state, what would now be called positive municipal law. The other two classes are not quite so clear, either in respect to the distinction between *ius gentium* and *ius naturale* or in respect to the relation of both to the *ius civile*. Cicero had used both these terms but had apparently made no distinction of meaning between them. In origin, as was said in the preceding chapter, the term *ius gentium* belonged to the lawyers, while *ius naturale* was a rendering of Greek philosophical terminology. In meaning the two apparently coalesced, both for the earlier lawyers and for Cicero. They signified indifferently principles that were generally recognized and therefore common to

the law of different peoples and also principles that were inherently reasonable and right without reference to their occurrence in any system of law. The distinction was easy to overlook because common consent was taken as a test of validity. It seemed a fair presumption that what many peoples have arrived at independently is more likely to be right than what is peculiar to any single people.

As time went on the lawyers apparently saw a reason for distinguishing *ius gentium* from *ius naturale*. Gaius, writing in the second century, continued to use the terms synonymously, but Ulpian and later writers in the third century made a distinction, as did also the lawyers who prepared the Institutes in the sixth. The distinction added precision to legal definition, but it perhaps signified also a more penetrating ethical criticism of the law ; even what is generally practiced may still be unjust and unreasonable. The main point upon which *ius gentium* and *ius naturale* are distinguished is slavery. By nature all men are born free and equal, but slavery is permitted according to the *ius gentium*. It is hard to tell just what this natural liberty meant to the lawyers who asserted its reality so flatly, but in view of the efforts made, not without success, to throw legal safeguards about slaves and other oppressed classes, it seems reasonably to construe it as representing some moral reservation about practices whose legality was unquestionable according to all known codes. Perhaps the idea was, as Professor Carlyle suggests, that in some purer or better form of society slavery had not existed, or would not exist. At all events, such passages would be so understood after Christianity had made the story of the fall of man a common belief.

Whether or not they distinguished between *ius gentium* and *ius naturale*, none of the lawyers doubted that there is a higher law than the enactments of any particular state. Like Cicero they conceived of the law as ultimately rational, universal, unchangeable, and divine, at least in respect to the main principles of right and justice. The Roman law, like the English common law, was only in small part a product of legislation. Hence the presumption was never made that law expresses nothing but the will of a competent legislative body, which is an idea of quite recent origin. It was assumed that "nature" sets certain norms which the positive law must live up to as best it can and that, as Cicero had believed, an "unlawful" statute simply is not law. Throughout the whole of the Middle Ages and well down into modern times the existence and the validity of such a higher law were taken for granted. As Sir Frederick Pollock says, the central idea of natural law, from the Roman Repubic to modern times, was "an ultimate principle of fitness with regard to the nature of man as a rational and social being, which is, or ought to be, the justification of every form of positive law."[7]

In theory, therefore, the positive law is an approximation to perfect justice and right ; these represent its objects and form its standards. It is, as Ulpian says, quoting Celsus, *ars boni et aequi.*

> Justice is a fixed and abiding disposition to give to every man his right. The percepts of the law are as follows : to live honorably, to injure no one, to give to every man his own. Jurisprudence is a knowledge of things human and divine, the science of the just and the unjust.[8]

Hence the lawyer is a "priest of justice," "the practitioner of a true philosophy, not a pretender to an imitation." It is not necessary to take Ulpian's rhetoric as a literal statement of fact. But it remains true that the Roman jurists did build up a more enlightened body of law than had ever existed, and though the

changes they wrought had their economic and political causes, it is not to be imagined that they came about without reference to the ideals of the profession.

Natural law meant interpretation in the light of such conceptions as equality before the law, faithfulness to engagements, fair dealing or equity, the superior importance of intent to mere words and formularies, the protection of dependents, and the recognition of claims based on blood relationship. Procedure was more and more freed from mere formality ; contracts were made to rest on agreement rather than on words of stipulation ; the father's absolute control over the property and persons of his children was broken ; married women became the full legal equals of their husbands in the control of their property and children ; and finally great progress was made in throwing legal safeguards about slaves, partly by way of protecting them against cruelty, partly by making their manumission as easy as possible. A modern exponent of "just law," Rudolf Stammler, has regarded this belief in justice as the crowning glory of Roman jurisprudence.

> This, in my opinion, is the universal significance of the classical Roman jurists ; this, their permanent worth. They had the courage to raise their glance from the ordinary questions of the day to the whole. And in reflecting on the narrow status of the particular case, they directed their thoughts to the guiding star of all law, namely the realization of justice in life.[9]

It should be noted that these reforms in the Roman law, though they were completed after the beginning of the Christian era, were not due to Christianity. The effective humanizing influence was Stoicism and there seems to be no evidence whatever of any effect of the Christian communities upon the great jurists of the second and third centuries. At a later date, in the time of Constantine and after, Christian influence can be seen, but it was not exerted in the directions mentioned above. Its purpose was to secure in one way or another the legal position of the church or of its officials, or to aid in carrying out policies of the church. Typical legal changes which the church secured for the protection of its interests were the right to receive property by will, the establishment of the jurisdiction of bishops' courts, the power to supervise charities, the repeal of the laws against celibacy, and the enactment of laws against heresy and apostasy.

Finally, the Roman law crystallized the theory, already contained in Cicero, that the authority of the ruler is derived from "the people." The theory was summed up in a sentence by Ulpian, repeatedly quoted, and there is no dissent by any of the lawyers either of the Digest or the Institutes :

> The will of the Emperor has the force of law, because by the passage of the *lex regia* the people transfers to him and vests in him all its own power and authority.[10]

The theory is to be understood, of course, in a strictly legal sense and it is couched in terms that had a definitely technical significance. In itself it justifies neither the implication of royal absolutism, which was sometimes derived from the first clause, nor of representative government, which the sovereignty of the people came to signify later. The latter meaning would have been especially absurd in the Roman Empire when Ulpian wrote. The idea behind Ulpian's statement is that expressed by Cicero, that law is the common possession of a people in its corporate capacity. This idea appears in the theory that customary law has the consent of the people, since custom exists only in the common practice. It

appears also in the classification of the sources from which law is derived. Thus law might arise by the enactment of a popular assembly (*leges*), or by the vote of some authorized part of the people such as the plebeian assembly (*plebescita*), or by a decree of the Senate (*senatus consulta*), or by a decree of the emperor (*constitutiones*), or by the edict of an ordinance-issuing official. In all cases, however, the source must be authorized and in the last resort all forms of law go back to the legal activity inherent in a politically organized people. In a sense every established organ of government does "represent" the people in some degree and some capacity, but there is obviously no implication that representation has anything to do with voting and still less that voting is a right inherent in every person. The "people" is an entity quite different from the persons who happen at any given time to be included in it.

At the same time some essence has been preserved from the ancient doctrine that law is an "impersonal reason" and that in consequence there is a broad moral distincion between lawful government and successful tyranny. Even though the former be often bad and the latter sometimes efficient, subjection to law is not imcompatible with moral freedom and dignity, while subjection to even the kindliest master is morally degrading. The Roman law preserved the spirit of Cicero's striking phrase : "We are servants of the law in order that we may be free."[11] And indeed, there is no more astonishing evidence of the strength that this conviction had come to have in European morals than the fact of its preservation in a system of law which reached its maturity at a time when the personal power of the emperors was often unlimited and when their authority rested frequently on nothing better than force. Yet the fact remains that in the long run the ideal embedded in the law was a permanent factor in European political civilization — a distillation from the old free life of the city-state — which was able to endure through and beyond an age in which all the servility of oriental despotism had apparently been transplanted to Rome. [pp. 163-173]

NOTES

1. George H. Sabine, *A History of Political Theory* (New York : Holt, Rinehart and Winston, 1937), pp. 163-173.
2. Cicero, *Republic*, 3, 22, trans., Sabine and Smith.
3. Cicero, *Laws*, trans., C.W. Keyes, 1, 10, 28-29.
4. A.J. Carlyle, *A History of Mediaeval Political Theory in the West*, vol. 1 (1903), p. 8.
5. Cicero, *Republic*, 1, 25.
6. Cicero, *Laws*, 3, 1, 2.
7. Sir Frederick Pollock, "The History of the Law of Nature," in *Essays in the Law* (1922), p. 31.
8. *Digest*, 1,1, 10.
9. Rudolf Stammler, *The Theory of Justice*, Eng. trans. (1925), p. 127.
10. *Digest* 1,4, 1.
11. *Pro Cluentio*, 53, 146.

Jurisprudence

Fifth Edition

R W M Dias

MA LL B **(Cantab.)**

of the Inner Temple, Barrister
Fellow of Magdalene College
University of Cambridge

The Reformation and the Renaissance to the Nineteenth Century

The result of the disintegration of the Holy Roman Empire was the replacement of the medieval idea of a united Christendom under the spiritual guidance of the Church by independent national states. The paramount need now was to free the national sovereigns from external authority and to consolidate power in their hands, since more and more strength was required to maintain independence in a world of increasing rivalry. Support for this movement came from various quarters. Individuals wished to free themselves

11 See pp 471-475 post.

from the Church and the shackles of the feudal system which had held Europe for so long in an iron grip, and in particular the rising commercial middle class wanted freedom to pursue its trade. The age of discovery and colonisation was at hand, and these enterprises were left to individuals, who had to be guaranteed liberty to carry on these activities. Individuals, therefore, found that a powerful sovereign was their best guarantee against interference, and the need was to foster the power of the sovereign.

Side by side with these developments natural law theory assumed a more secular aspect, brought about by a combination of factors. It was necessary to adapt theory to meet the above needs. Advances were also being made in knowledge about the physical operations of the material world. The Thomist system contained no scientific treatment of the movement and measurement of bodies and forces, and its exponents found themselves intellectually incapable of assimilating the new knowledge without some adjustment. The effect was not a rejection, but an adaptation of natural law theory. The Protestants denied that the Church had authority to expound the law of God, and in the controversy appeal to the law of nature as based on the reason of the individual came more and more into prominence. The intellectual world came to be actively engaged in the revived study of Roman law, which was now being adapted to meet contemporary local needs. This had two results. The diffusion of Roman law throughout different nationalities in Europe led to an association of it with natural law. It was revered as being supremely reasonable, the spirit of Rome still ruled the world by her reason. In this way the resources of Roman law were always on tap to supply the deficiencies of positive law; it also provided most of the material for constructing a system of international law. Secondly, the adaptation of Roman law had led to local variations. This fostered the belief that there was a 'natural law' which transcended its innumerable applications. The result of this belief was that the study of 'pure' Roman law became mainly an academic discipline[12].

Reflecting these developments and the call of a new age for sovereign power, theory tended to develop in the direction of justifying absolute monarchy. In this connection one thinks immediately of Machiavelli's *Prince* (1513), but his ideas are interesting chiefly on account of what he said in his other work known as *Discourses*. He characterised republics as being 'free' and superior to princedoms in that a high degree of virtue is required of people not to abuse freedom. To the extent that people lack virtue, they are 'corrupt' and have to be governed by a tyrant or prince, and in his *Prince* he elaborated absolute dictatorship. The state and its sovereign have to be supreme and subject to no external control, while moral and ethical principles need to be subordinated to political expediency.

In the hands of Thomas Hobbes (1588–1679), described as a belated medievalist, absolute sovereign power was justified by postulating an imaginary compact between ruler and ruled. In a state of nature everyone is entitled to everything; but this leads to friction. So Hobbes believed that Man's life in a state of nature was one of fear and selfishness. 'The life of man', in his own words, was 'solitary, poor, nasty, brutish and short'[13]. It is a fundamental dictate of the law of nature that people desire peace[14], so to

12 Cf Savigny's attitude: ch 18 post.
13 *Leviathan* ch xiii.
14 *Leviathan* ch xiv.

escape from the state of nature they entered into the social contract, whereby each surrendered his rights to a sovereign ruler, the Leviathan, who, in return for absolute subservience, guaranteed peace and greater security to each than he might otherwise have had. In this way natural law theory had come to support power[15]. It is important to remember that Hobbes lived through the Civil War in England, so his preoccupation was with stable and secure government. His doctrine reflects the idea that the security of the individual can only be found in the exaltation of the power of the sovereign. Liberty is not of men, but of the sovereign[16]. The menace to the individual from the sovereign is not faced. There is a suggestion that the sovereign is bound by natural law; but this is nothing more than a vague morality. The sovereign is expected to allow a harmless liberty, but there is no justification for resisting laws. There is no such thing as an iniquitous law; whatever is law is moral. Since society depends on the sovereign, law in society is what the sovereign commands: form rather than content becomes the determining feature of law. Although Hobbes was not concerned with protecting the individual against oppression by the sovereign, it was of the utmost significance that one precept of his natural law was the individual's entitlement to self-preservation. For herein lay the germ of a new view of natural law as a system of 'rights' in the individual. It is a far cry from the ancient idea of it as laying down duties. In Hobbes's scheme the authority of the Church as the interpreter of God's law is vigorously denied. All power is given to a severely utilitarian secular sovereign. This marked the final break up of the Catholic international order of things.

In this way national sovereigns acquired freedom from external power, but this was taken to imply complete liberty of action in their mutual dealings. Side by side with this they acquired also unfettered power over their own subjects. It was not long before the need arose to curb the abuse of liberties in international relations and the abuse of domestic power over subjects.

It has been pointed out that unlimited freedom to act as one pleases leads ultimately to chaos; and so it proved in the international sphere. Under the guise of the 'sovereignty of states' the unbridled pursuit of selfish policies reduced Europe to a state of barbarity culminating in the Thirty Years War. This abuse of liberty evoked increasingly loud cries for some power-structure whereby such freedom of action might be restrained. Voices, notably that of Hugo Grotius (1583–1645) started to preach the need for restraint[17], and a body of duties, known later as 'international law', began to be worked out in the hope of setting limits to the conduct of states. In his natural law theory Grotius was essentially a child of the time, but a feature of interest and significance in his writings is the shift in emphasis from Divine Providence to the reason of Man. He believed that natural law was rooted in the nature of Man, and would exist even if there were no God. This was not a denial of the Deity, but an assertion that natural law is independent of God and a quality of Man. The binding force of law is its essence. The problem is how men, supposed to be free, could fetter themselves. The answer was: by voluntary submission, by a social contract. Mutual promises have no obligatory

15 *Leviathan* chs xvii–xviii. Cf the Jewish social contract, which was in support of immunity from power.
16 *Leviathan* ch xxi.
17 *De Jure Belli ac Pacis*, *Prolegomena* 28.

force unless it derives from some rule giving such force to promises; and this rule is a rule of natural law. By virtue of it Grotius sought to explain the adoption of different forms of government by different societies. Reason impels Man to seek society. The state originates in a contract by virtue of which each individual surrenders his sovereignty to a ruler. The group has the liberty to choose the order which it prefers, but once it has chosen it loses all power to restrain its ruler. This was the way in which absolute power emerged. The subjects are bound to an almost blind obedience and only in exceptional circumstances did Grotius concede a liberty to revolt. The only guarantee which they have against an abuse of power by the ruler is that he is bound by natural law, which Grotius explained in these words:

> 'A dictate of right reason, which points out that an act, according as it is or is not in conformity with rational nature, has in it a quality of moral baseness or moral necessity.'

The next, and more important, step was to restrain the absolute sovereign in his relations with other such sovereigns, for it was the absence of restraints that precipitated wars. Grotius accordingly went on to assert that nations were still in a state of nature and so bound by natural law. Man according to its fiat is a sociable being and desires peaceful society. For nations to assert their absolute independence was futile, and the sooner they realised that they, too, should unite to form a society of states the better it would be. It is these doctrines that have earned for Grotius the distinction of being styled the 'father of International Law'. He sought to establish international law on two foundations, the law of nature and consent. From natural law were deduced principles as to how states should conduct themselves, the idea of consent was utilised to infer rules from the observed practice of states. The difficulty about the latter is that one cannot derive an 'ought' from an 'is'; in other words, in view of the variety of reasons why different states at different times behave in a given way it is futile to propound a norm of behaviour. However, Grotius posited natural law as the principal foundation of international law.

In the municipal sphere the individual, who so fondly trusted his sovereign to deliver him from feudal oppression, soon found that he had only exchanged one tyranny for a worse. This gave rise to domestic struggles for immunity from the abuse of the sovereign's power. The age of the individual was now at hand, an age of commercial adventure and expansion, an age of trading middle classes. In support of this movement natural law doctrine was again called in aid. As early as 1581, when the Netherlands broke away from the overlordship of Philip II of Spain, it was declared in the Act of Abjuration that if a sovereign treats his subjects like slaves he ceases to be their rightful sovereign. In the succeeding age the mythical social contract theory was re-furbished by John Locke (1632–1704)[18]. The reign of the Stuarts in England had revealed what a menace the sovereign could be to the individual. So Locke evolved a theory to protect the individual. The state of nature, which preceded the social contract, was not one of anarchy, as Hobbes had imagined, but was 'a state of liberty, not of license'. It had, however, one defect, namely that 'property' was insecure and by 'property' was meant life, liberty and estate[19]. 'Every man has a *Property* in his own

18 Locke *Second Treatise on Civil Government* (1690).
19 *Locke* para 123.

Person', wrote Locke; and since the source of value lies in a man's labour, whatever a man 'hath mixed his *Labour* with, and joyned to it something that is his own' becomes his[20]. 'Property' in the sense described was insecure because (a) there was no established law, nor (b) an impartial judge, and (c) the natural power to execute natural law was not always commensurate with the claim. Executive powers were regarded as natural powers of a quasi-judicial character. To remedy this flaw Man entered into the social contract by which he yielded to the sovereign, not all his rights, but only the power to preserve order and enforce the law of nature[1]. The individual retained the 'natural rights' to life, liberty and estate, for they were the natural and 'inalienable rights of Man'[2]. This was an innovation of significance inasmuch as the idea of a natural law of society was transmuted into one of individual 'natural rights' which could be used to resist social control. The purpose of government was to protect these; it has no other end than 'to preserve the members of that society in their lives, liberties and possessions'. Its function is in the nature of a trust[3]; as Locke put it, all this 'is "*limited to the publick good* of the society"'[4]. So long as government fulfils this purpose, its laws should be binding. When it ceases to protect or begins to encroach on these 'natural rights', laws lose their validity and the government may be overthrown. In this way Locke championed the revolution of 1688–1689, and his idea that positive law might thus be overborne by natural law sustained the American colonists in their successful defiance of the British Parliament in the fateful years 1775–1781.

It is clear that in Locke's view unlimited sovereignty is contrary to natural law. The basic limitations are (a) that sovereign power cannot justifiably be used in an arbitrary fashion over the lives and fortunes of people. No individual enjoyed such power in the natural state and none could therefore be surrendered to the sovereign. (b) The sovereign may not rule by arbitrary decree, but only through known laws administered by known judges. (c) The people may not be deprived of their possessions without their consent. (d) The sovereign may not transfer the power of making laws to other persons. In order that the sovereign shall not abuse his authority Locke championed a constitutionally limited sovereign and a threefold division of governmental power[5]. Foremost there stands the legislative power for the creation of rules to give effect to and protect the 'inalienable rights'; secondly, there is the executive power by which the law is enforced; and thirdly, there is the 'federative' power, which concerns the making of war and peace and controls the external relations of the state. Locke thought it advisable to confer the legislative and executive powers on different organs, one reason being that the legislature does not, while the executive does, operate continuously, and another being the danger of entrusting law-makers with the power to carry out the laws which they themselves make. He did not advocate a similar separation between the executive and 'federative' powers since both are dependent on force. His views had considerable influence, for the climate of opinion was such as to make them acceptable. Not only was he hailed as the

20 *Locke* para 27.
1 Cf Pratt CJ: 'The great end for which men entered into society was to secure their property': *Entick v Carrington* (1765) 19 State Tr 1029 at 1066. Cf *Locke* para 124.
2 *Locke* para 135.
3 The idea of a trust was echoed in *Jilani v Government of Punjab* Pak LD (1972) SC 139 at 182.
4 *Locke* para 135.
5 *Locke* ch xii.

intellectual defender of the English constitutional revolution of 1688–1689, but he also influenced the later American Revolution.

The problem of protecting the individual was not to be solved simply by curbing or destroying the sovereign, for power was only transferred from him to a faceless institution known as 'government'. One way of controlling power, whether exercised by sovereign or government, is to postulate the supremacy of law. Long before Locke, Chief Justice Coke (1552–1633) maintained this, and his view of law was essentially naturalistic. Law commands obedience, he said, by virtue of being reasonable; and it is in this context that his doctrine of the avoidance of even Acts of Parliament, which offend against common right and reason, has to be understood[6]. He also introduced an idea of profound significance when, in his famous disputation with James I, he insisted that even the king is under God and the law, and added that the reason of the law is not 'natural reason', but 'artificial perfection of reason, gotten by long study'. The implication of this is that kings and governments are controlled by the very special craftsmanship of the law of which judges, by virtue of their experience and practice, are the exponents. This made the judges the champions of the individual, and stresses the importance of an independent judiciary in controlling power.

Another way of controlling governmental power was put forward by the French philosopher Montesquieu (1689–1755), who purported to base his conclusions on the English constitution. The secret of its success, as he put it, lay in the fact that power was not concentrated in the hands of an individual or group. Unless power is made to check power, it will be abused. An essential prerequisite is that general rules should be laid down in advance of actual situations, but even so there remains the real possibility of abuse if the person laying down the general rules were also allowed to apply them and then carry out his own decisions. So the guarantee against abuse lies in establishing mutual checks between the legislative, executive and judicial functions[7]. All that his doctrine of the 'separation of powers' insists on is the danger of an absolute union of these functions in one person or body. It does not require their respective allocation to three distinct organs with no inter-relation between them. The functions may well be shared as in England, for instance, where the cabinet, which is the executive centre, is an integral part of the legislature and responsible to it; executive departments have legislative functions entrusted to them; justices of the peace perform judicial and executive functions. Montesquieu's doctrine is carried furthest in the United States of America, where the separation is guaranteed by a Constitution, guarded by the courts. The federal legislative power is vested in Congress, the federal executive power in the President and his cabinet; and the federal judicial power in the Supreme Court. A very great deal of inter-relation exists, which enables them to work harmoniously

In France the continuing need to protect the individual against an oppressive monarchy found expression in the theory of Rousseau (1712–1778), in which the idea of the social contract underwent yet another revision. Rousseau set out to evolve a community in which the community as such

6 *Bonham's Case* (1610) 8 Co Rep 114a.

7 Montesquieu *L'Esprit des Lois* (1748) 11, 4–6. It is commonly supposed that Montesquieu based his views on the British system. For the view that he was really arguing what ought to obtain in France and that he used Britain as a façade for avoiding domestic censorship, see Merry *Montesquieu's System of Natural Government* (1970).

would protect the individual; but in which at the same time the individual would remain free from oppression. All should participate in policy-making. Accordingly, he argued that in the original contract the individuals did not surrender their rights to any single sovereign, but to society as a whole, and this is their guarantee of freedom and equality. Society, having come into being for this purpose, is expected to restore these rights to its members as civil liberties. Their basis is a moral one. Each individual is not subject to any other individual, but to the 'general will'[8], and to obey this is to obey himself. Government and law are both dependent upon general will, on popular as distinct from parliamentary sovereignty, which may revoke or overthrow them. Enacted law is necessary to avoid arbitrariness, and to get law right to begin with public servants should apply the 'general will' so that law will produce just results[9]. Little wonder that this theory was utilised as the philosophy of the French Revolution in order to get rid of the monarch. Yet again the pattern of history was repeated. Deliverance from the evil of uncontrolled sovereign power only led to the evil of uncontrolled popular liberties as manifested in the excesses of the Revolution; which in turn called for a return to power in Napoleon.

The same sad swing can be detected in Germany[10]. The nineteenth century saw the unification of the Germanic states, which stimulated an upsurge of nationalistic fervour. Although there was some talk of a 'Law State' (*Rechtsstaat*, constitutional government), the trend was towards the glorification of imperial power. The reaction against this after World War I came with the Weimar Republic, which was a 'national-social state'. On the one hand, its socialist character required a planned economy, which meant concentrating power in these matters in government. On the other hand, the extremely permissive attitude adopted by government in its reaction against the authoritarianism of the past allowed disruptive and revolutionary freedom to flourish to the point of eventually reducing the country to economic and social chaos. Little wonder that there was general support for the return to a power-structure under Adolf Hitler to restore stability. There is no need to dwell on the way in which that power was subsequently abused[11]. After World War II West Germany became a 'republican, democratic and social Law State'[12]. Its socialist character requires concentration of power in government in economic planning, but at the same time, as with the Weimar Republic, extreme sensitivity to the excesses of the Hitler régime has induced a permissive attitude towards disruptive movements.

The Soviet Union provides a vivid example of the power-freedom-power swing. The Revolution was a successful bid for freedom from the power of the Czars. In its wake the revolutionary leaders immediately claimed for themselves unlimited freedom to act as they chose, untrammelled by codes or laws[13]. More significantly, they have established an infinitely more ruthless power-structure, rooted in a new ruling class of the Communist Party, which brooks no opposition in any shape or form. This was required, among other things, to stifle all contrary values, which are condemned as 'counter-

8 Rousseau *The Social Contract* (1762, Everyman's Lib) pp 13–15.
9 *Rousseau* pp 26–27.
10 See generally Dietze *Two Concepts of the Rule of Law*.
11 See p 387 post.
12 Bonn Basic Law, art 28.
13 See ch 19 post.

revolutionary', and also to carry out a long-term programme of mass re-education of the populace in the values of communism. The lengths to which the Soviet Union is prepared to carry this power is seen in the way in which she intervened to crush libertarian stirrings in Hungary in 1956 and in Czechoslovakia in 1968.

There have been other power-structures which have filled many people all over the world with misgiving. Traumatic experiences such as these have induced at times an almost hysterical reaction against power and, indeed, any call for law and order. It has the unfortunate result of blinding people to the alternative dangers of the abuse of liberty through overweening permissiveness. The need to prevent this abuse is at least as great as the need to prevent the abuse of power. Democracy was a device to prevent the latter. Even among a small group of people, as long as the person in authority enjoys their trust, they will be content to leave power in his hands. Once he loses it through wickedness or mismanagement, they will feel driven to demand a say in their own government, either personally in the case of very small groups, or through representatives in the case of large groups. Representation dilutes democracy, unavoidably in the nature of things, and it leaves democratic processes of election and decision open to manipulation by those in quest of power. By means of propaganda and by taking advantage of the apathy of the majority, who do not attend meetings, or by shouting down opposition, they manage to vote themselves on to bodies, which then use the votes of these few as block-votes to represent the whole constituency. Having won power in this way, they seldom risk losing it by allowing any reference back to the constituency on a one man one vote basis. Although such processes as these are claimed to be democratic, they are a perversion of democracy. Apathy may deserve what it gets, but election through the apathy of the majority is not election *by* the majority.

Democracy is workable as long as there is a substantial area of shared values and aspirations among the people and where they have the maturity to rise above differences[14]. Once this willingness to 'rub along together' disappears and differences become irreconcileable, (eg with racial or religious divisions), and fragmentation reaches the point of pluralism without a cohesive substrate, then government through democratic participation becomes precarious, if not impossible. For it may then not be possible to reach majority decisions, and, even if it were, one section or other may refuse to abide by them[15]. The uncomfortable question arises whether democracy itself is only an intermediate stage in the doom of society, beginning with loss of confidence in a tyrant and ending in anarchy. The answer is that it need not be. The message coming through history is that to break out of the melancholy spiral of power and liberty a way must be found of curbing both rather than promoting either[16]. This, it would seem, is the paramount need of the present age in its quest for justice.

14 Cf Machiavelli *Discourses*, and p 79 ante.
15 Eg Northern Ireland.
16 Lord MacDermott *Protection from Power under English Law* p 195.

TRANSCENDENTAL IDEALISM

The close of this epoch was marked by another line of thought, which was equally uncompromising in its insistence on individual freedom, but wholly idealistic in character. Kant (1724–1804), whose doctrines were developed by Fichte, taught that sensory perception is the avenue to knowledge of the objective world, but that all such perception is shaped by *a priori* preconceptions. Thus, preconceptions of space, time and causation filter one's experience of nature. In so far as Man is part of the world of reality, he is subject to its laws and is to that extent unfree, but his reason and inner consciousness make him a free moral agent. Man thus participates in two worlds, the 'sensible' and the 'intelligible'. Law and morality belong to the latter. The actions of Man as a free agent are governed by aims, and the ethical basis of action has also to be accepted *a priori*. Justice, according to Kant, originates in pure practical reason. People know *a priori* how to act justly. The ultimate aim of the individual should be a life of free will; but it is when free will is exercised according to reason and uncontaminated by emotion that free-willing individuals can live together. People are morally free when they are able to obey or disobey a moral law. Kant propounded two principles of practical reason. (a) 'Act in such a way that the maxim of your action can be made the maxim of an universal law (general action)'[3]. This is his famous 'categorical imperative'. (b) 'An action is right only if it can co-exist with each and every man's free will according to universal law'. This is his 'principle of right'[4].

2 See pp 79–84 ante.
3 Kant *Philosophy of Law* (trans Hastie) p 34.
4 *Kant* p 45.

The first point to note is the emphasis on the individual. A feature of Kant's doctrine is his proclamation of the autonomy of reason and will. Human reason is law-creating and constitutes moral law. Freedom in law means freedom from arbitrary subjection to another, and law is the complex totality of conditions under which maximum freedom is possible for all. To this end a separation of powers is necessary to prevent the emergence of a despotic régime, and the sole function of the state is to ensure the observance of law. Kant proceeded to urge that the individual should not allow himself to be made the means to an end, since he is an end in himself[5]; and that he should, if need be, retire from society if his free will would involve him in wrongdoing[6]; for Kant did perceive the necessity for rules in social existence, guided by a just general policy. Society unregulated by right results in violence. Social existence and violence are incompatible; so reason demands that Man has an obligation to enter into society and to avoid wronging others. Such a society has to be regulated by compulsory laws, and if these laws are derived by pure reason from the whole idea of social union under law, Man will be able to live in peace[7]. What is needed is a rule of law, not of men[8]. The second point is that the Kantian ideal of laws bears no relation to any actual system of law. It is purely an ideal to serve as a standard of comparison, not as a criterion of the validity of law.

[1]

Social Philosophy & Policy Vol. 4 Issue 1 ISBN 0625-0525

TAXATION IN A LOCKEAN WORLD*

BY RICHARD A. EPSTEIN

'Tis true governments cannot be supported without great charge, and it is fit everyone who enjoys a share of the protection should pay out of his estate his proportion for the maintenance of it. But still it must be with his own consent, i.e., the consent of the majority giving it either by themselves or their representatives chosen by them. For if any one shall claim a power to lay and levy taxes on the people, by his own authority, and without such consent of the people, he thereby invades the fundamental law of property and subverts the end of government. For what property have I in that which another may by right take when he pleases to himself.[1]

I. LIBERTY AND TAXATION

One constant refrain of political and constitutional history treats taxation as an inherent and indispensable power of the sovereign. In one sense that proposition seems to be a self-evident truth: if the sovereign is without resources, then it cannot carry out any mission, however defined. It must atrophy and die. Taxes are to the state as food is to the body. Yet in another sense the proposition is one that leads to anxiety and disquiet. *Pace* Locke, the institutions of representative government do not work as smoothly as we might hope. Taxation is the power to coerce other individuals to surrender their property *without* their consent. In a world – a Lockean world – in which liberty is regarded as a good and coercion an evil, then taxation authorizes the sovereign to commit acts of aggression against the very citizens it is supposed to protect. The two sides of the dilemma are thus complete. A world without taxation is a world without government; a world with taxation is a world with institutionalized coercion, where the citizen is made the servant instead of the master of the sovereign. Liberty and

* An earlier version of this paper was presented at a conference on "Taxation and Liberty" organized by Alvin Rabushka and held in Santa Fe, New Mexico in September 1985, at the law and economics workshop at the University of Toronto earlier that month, and at the Washington University Business School in January. I should like to thank Walter Blum, James Buchanan, Joseph Isenbergh and William Niskanen for their valuable comments on earlier drafts of this paper.

[1] John Locke, Of Civil Government (Second Treatise) ¶140, 142.

sovereignty are in conflict. With the tension unresolved, we are left only with the unhappy choice between anarchy and totalitarianism.

One central question for any political theory is how to escape the horns of this dilemma. More precisely, the question is how to preserve the power of taxation while curbing its abuse. The most promising line of inquiry, I believe, is to show that taxation, while an inherent power of the sovereign, is a *limited* one as well. This involves applying the general principles of limited government to the program of taxation. This effort to find that middle course invites a set of inquiries that are avoided with both extreme positions. What are the limitations upon the sovereign power to tax? And how do they preserve the liberties of the citizens so taxed? That question itself has two separate parts: first, what is the *substantive content* of the limitations that should be imposed and, second, what are the *institutional arrangements* that give these limitations a fair chance of success?

The failure to answer either of these questions in definitive form has led to the modern consensus that veers too far in the direction of untrammeled state power to tax. The first part of the conventional wisdom is that democratic political institutions are the prime source of protection against abuse of the power of taxation. The risk of exploitation is reduced because no single individual is given the power to tax without being accountable to the public at large, which can voice its collective objection to oppressive taxes at the polls, a check to which all politicians pay homage. Within the American system at least, however, the second part of the consensus answer specifies only weak constitutional restrictions upon the power to tax. Taxes can be struck down if they rely upon a suspect class or limit a preferred individual freedom. An income tax of 10 percent on whites and 20 percent on blacks, or vice versa, is invalid because it relies upon a racial classification. A special onerous tax upon magazines and newspapers is invalid as an infringement of the freedom of speech. These limitations upon the power to tax still leave the legislature enormous discretion with respect to the objects and the level of taxation; and they have done nothing to limit the vast proliferation of special provisions in the Internal Revenue Code that have proved remarkably resistant to intelligent reform nor, in earlier days, to question a steeply progressive tax with top marginal rates that nominally exceeded ninety percent. A clear theory of individual rights and limited government points to more robust constitutional limitations upon the power to tax than the political consensus now accepts.

Both the objects of taxation and the rates are covered by my analysis, which links the modern themes of tax equity to the general theory of political obligation. The paper proceeds as follows. Section II outlines the basic political justification for taxation. Section III explains the basic taxation bargain that this theory calls forth. Section IV then shows how that bargain

can be applied to technical questions of both the tax base and the rate structure. Section V gives a brief account of the possible constitutional limitations upon the government's power to tax.

II. THE ORIGINAL POSITION

The American tradition of government has been heavily influenced by Lockean social contract theory. The central question for Locke is how one is to move from a world with a large number of separate and autonomous individuals into a world with a single sovereign able to maintain political order, all the while protecting both liberty and property against government domination. In the Lockean view, all individuals are endowed with rights of personal liberty and rights to acquire property,[2] rights that control against theft, aggression, and fraud. Yet in the absence of any state the only remedy available to these individuals is self-help, whereby force resists or combats force. The weakness in this system lies not in its basic conception of entitlements. Locke himself accepts the set of entitlements that has in fact shaped both the civil and criminal law of virtually all civilized societies. I know, for example, of no one who defends murder, rape, or theft. Similarly, no one thinks that the right of self-defense should be abolished as a matter of first principle, even if they think that the existence of an organized police force calls for pragmatic restrictions upon its use.

The difficulties of the Lockean view lie in the vagaries of its enforcement. It is one thing to believe in the right of self-defense. It is quite another to determine who is the aggressor and who is the victim. To allow every person to make that determination for himself invites in private affairs what is so common in international relations: aggression under the pretext of defense. Similarly, there are limits to the right of self-defense even when the use of force is justified: once the peril has ceased the further use of force is excessive and outside the boundaries of the original justification. But, again, these limits are neither self-defining nor self-enforcing.

[2] There are difficulties in Locke's account of the original position, which treats all external things as being owned in common while each individual has exclusive ownership of his own person. Locke himself thought that his labor theory of value explained why any individual could take land or other property from the original common pool. But his theory is subject to the criticism that the private ownership it creates is subject to a lien for the benefit of the public at large, for otherwise the individual acquirer gets more than his labor has added. Locke's case would have been made more forcefully if he had taken the basic position of both the common and the Roman law – that all external things are unowned in the original position – for then first possession need not defeat any prior claims. The original no-ownership rule has a utilitarian simplicity, for there is little need to worry about a monopoly over natural resources when each person owns his own labor. For a defense of the common law solution, see Richard A. Epstein, Takings: Private Property and the Power of Eminent Domain, Chs. 2 & 20; and *Possession as the Root of Title*, 13 Ga. L. Rev. 1221 (1979).

The state offers a practical way to escape these enforcement difficulties by providing a single umpire whose decisions bind both contestants. Adjudication and public force replace self-help. Difficulties beset the implementation of the plan by the formation of government. Sovereignty sounds like an abstraction, but the power is always in the hands of natural persons ("real people") who may use their monopoly of force as a cloak for aggression against the very persons whose liberty and property they are sworn to protect. *Quis custodiet custodies?*, "Who guards the guardians?" is still the most probing question ever asked in political theory. In light of the manifold disasters of this and every other century, one could almost wish for a world without organized government on the commendable theory that random assaults offer a smaller risk than organized terror. But that solution is simply not in the cards. The void created by the lack of central authority will be filled by some desperado who will then disarm all his rivals. The practical choice therefore is not the choice between government and no government; the consistent anarchist will lose all (hypothetical) wars to the determined statist. The only set of feasible choices therefore are on the second order questions: what kind of state, with what kind of governors, minimizes the abuse of centralized power?

The options are limited and the risks are high, but on balance the best answers seem to come from the cross between the natural rights and the utilitarian tradition that together have inspired contract theories of the state and constitutional theories of limited government. Ideally, one can conceive of a (hypothetical) bargain calling for the universal renunciation of force by all persons, present and future. That agreement promises enormous gains to all the parties to it.[3] The theory of the state fleshes out the terms of that agreement.

The agreement is not real, so how can the argument proceed? Certain individuals might think themselves better off in a world that does tolerate the war of all against all, because they think that this is one war they can win. One person, therefore, might wish to hold out against the multitude. The greater the number of persons, the more likely that one of them will respond in this way. The theory of natural rights is designed to counter just this possibility by noting that the *baseline* against which "being better off" is measured gives to each person only his natural rights to liberty and property, and not the fruits acquired by theft, aggression, or deception.[4] The impulse

[3] The metaphor of bargain has been used to describe other situations in which coercive forced exchanges are designed to overcome common pool problems. Bankruptcy is one such example; see, e.g., Jackson, *Bankruptcy, Non-Bankruptcy Entitlements, and the Creditors' Bargain*, 91 Yale L.J. 857 (1982).

[4] See Frank Michelman, *Ethics, Economics, and The Law of Property*, 24 NOMOS, Ethics, Economics and the Law 3 (1982) for a vision of the world without such a baseline; and the brief rejoinder by Harold Demsetz, *Professor Michelman's Unnecessary And Futile Search For The Philosopher's Touchstone*; *id.* 41.

behind that intuition comes from two sources: the first is the self-evident belief that individual autonomy is a primary and indisputable good, and the second is that the overall level of human satisfaction and happiness is greater in a world without aggression than in a world with it.

Even if the baseline can be set as Locke would set it, the operation of the state cannot rest on actual or tacit consent. Instead, the justification for the state depends upon the *network of forced hypothetical exchanges* needed to overcome the persistent problems of collective action: holdouts and free-riders flourish in a world of high transaction costs.[5] Armed with that (inchoate) realization, the Lockean solution was simple and correct. It postulated a world in which, as a first approximation, every individual was required to surrender his right to use force to the state in exchange for the like promise of every other individual to so refrain. In such a world, the gains from being free of risk of aggression are worth more than the unqualified rights of self-help they replace. The risk of aggression of course remains, and self-help (now subject to judicial scrutiny and review) remains for emergencies where state intervention cannot be timely. Starting from the original set of entitlements, the formation of the state looks very much like a Pareto-superior move in which some natural rights are surrendered in exchange for the greater protection of the natural rights that are retained. Taxation is the form of surrender, and security is the good received in exchange. In the modern language of welfare theory, one distribution of rights is Pareto-superior to another so long as one person is better off in the first world than in the second, and no one else is worse off.[6] The formation of the state appears to satisfy this social welfare condition. Indeed, one can go further and argue that a more stringent requirement is satisfied as well. The gains of organized society cannot be captured solely by a small few. State protection is general so that each person shares in the social gains, roughly pro rata, in accordance to the value of his original endowments. Insisting that the surplus be divided in this unique way increases the definiteness of legal rights, which in turn reduces the danger of wealth loss from factional struggle.[7] In principle, no person can object to the second world in which the use of sovereign power leaves him better off than he was with his natural endowments. Similarly, no person can be allowed to stand outside the compact, for then the entire arrangement will unravel; for without universal compliance, the monopoly of the state is lost and the regime of self-help,

[5] See, generally, Russell Hardin, Collective Action (1982) for an account of these difficulties.

[6] See, e.g., Coleman, *The Economic Analysis of Law*, 24 NOMOS, Ethics, Economics and Law, Ch. 4 (J. Pennock & J. Chapman eds. 1982).

[7] In contrast, note that a per capita division of the gain from public expenditures may require explicit cash transfers between members of the society. Who is to pay and who is to collect is far from evident. The pro rata position is designed to avoid the social losses that would follow from any effort to redistribute the surplus by direct transfer payments.

with all its deficiencies, will emerge anew. The reconciliation of liberty and coercion is found in the stipulation for a return benefit, unobtainable through voluntary agreement, that exceeds in value the losses of liberty and property entailed by the formation of the state.

III. THE TAXATION BARGAIN

The task of governance does not end with a simple statement of the theory that justifies sovereign power. A Lockean theory leaves the state with functions to discharge – the keeping of order. The state must have the resources to discharge them. These funds could in principle come from voluntary contributions, but that prospect is bleak. Just as voluntary agreements could not form the state, so they cannot fund its continued operations; free riding endures. Everyone has the incentive to let other persons pay the private costs of the public good. The system is then in fundamental disequilibrium, for the state must have the resources it cannot command. Taxation from all is the only viable source of revenue to restore the equilibrium implied by the simple proposition that the state must have a secure revenue base. The question is distribution and incidence of the tax in question. At this point, one might expect that Locke himself would discuss in some detail the nature and the sources of the tax, but apart from his keen sense that taxation should be proportionate to the property protected,[8] he veers away from the issue.

Yet so great is the power of his theory of the state that it does more than account for the existence of the power of taxation. It also provides a blueprint for the closer analysis of the particular rules that limit the exercise of the taxing power. That such limitations are needed seems evident given the constant temptations for self-aggrandisement in politics. Even if all individuals abandon the use of force, they still remain political actors able to resort to influence and intrigue. The modern theory of public choice makes it abundantly clear what everyone who has heard of factions has known intuitively all along. There is a constant pressure for special interest legislation that is fundamentally at odds with the Pareto-superior pattern of forced exchanges appropriate for state initiation. In its place are efforts to achieve implicit transfers of property and wealth from one group to another. In the effort to skew the gains from collective life, taxation can be, and has been, used as a weapon of confiscation.[9] The Lockean theory of government calls for the use of coercion only to create a larger pie with all persons sharing in the gain. In practice, resort to governmental power often leads to a

[8] See note 1, *supra*.

[9] See Epstein, *Taxation, Regulation and Confiscation*, 20 Osgoode Hall L.J. 433 (1982).

smaller pie. Some individuals will obtain larger slices; otherwise nothing will drive the process. But others necessarily will be left with far less than they had before the government program was placed in operation. The standard Pareto measures are systematically violated.[10] There is no possibility that all will be allowed to share pro rata in the social gain. Democratic processes may make it more difficult to form a winning coalition; and they are certainly superior to despotic rule. But they do provide a new arena for abuse that cries out for institutional control. What form should the taxation bargain take?

A. Tax neutrality. A sound tax is designed to interfere as little as possible with the ordinary decisions that individuals would make in the investment and consumption of their capital and labor. On the Lockean theory, the sphere of individual autonomy should be invaded as little as possible in order to secure the maintenance of public order. In the original state of nature no individual can pass judgment on what others decide to do with their liberty or property unless the use of force or fraud is involved. The creation of a system of government should strive not to reduce the scope of permissible individual choice from what it was before. Accordingly, no person or group should be able to use the tax system to change the pattern of preferences of other individuals. If A ranks a set of (noncoercive) alternatives 1 to n before taxation, then A's ranking of those alternatives should remain 1 to n, without alteration, after the tax is imposed. So, too, with all other persons. If subsidies and penalties are generally forbidden, then they are forbidden not only by direct regulation, but by taxation as well.

The postulate of tax neutrality immediately follows.[11] Tax neutrality is a principle that recognizes that before a system of taxation is imposed, individuals order their preferences in a discrete and particular way. The ideal of tax neutrality simply provides that the system of taxation, as far as possible, should preserve the relative priorities that individuals attach to various activities. The function of the state is to protect liberty and property. It is not to aid one group or another in skewing the uses to which individuals put their natural endowments. Observing the principle does more than keep the relations between taxpayer activities constant. It also keeps their relative wealth constant as well, so that an increase for one is matched by a

[10] I explore some of these tensions in Richard A. Epstein, Takings, at Ch. 2, 12, and 18; *Toward a Revitalization of the Contract Clause*, 51 U. Chi. L. Rev. 703 (1984); and *Why Restrain Alienation?*, 85 Colum. L. Rev. 770 (1985).

[11] Here tax neutrality must be kept distinct from the idea of "revenue neutrality" that has been prominent in recent discussions of tax reforms. Revenue neutrality provides that the total revenue collected by the state should not be decreased by any substantive principles of tax reform. But it does not tell us how to decide the soundness of one basic structure against another. As a matter of first principle, once the right structure is worked out, the tax rates should be set so as to provide funds sufficient to discharge the public obligations.

proportionate increase for others. If tax neutrality could be perfectly achieved, the laws would act as a prism which magnified equally all preexisting endowments. The nature of private activities would not change, nor would the relative endowments of private persons.

The desirability of tax neutrality also follows from the more overtly economic view of the subject. In a well functioning economy, every actor should have to face the full opportunity costs of his private decisions. Where private rights in labor and property are well-defined, that normally will be the case. A system of taxation that imposes differential taxes necessarily subsidizes some activities and penalizes others, thereby distorting relative prices and, in consequence, private decisions on consumption and investment. Likewise, if taxation were allowed to make some persons better off and others worse off, struggles would arise between persons who would want to funnel wealth through the tax system and those who would want to keep it out of the system. These resource losses would reduce the likelihood that any tax imposed will be Pareto-superior, for some persons could and would emerge net losers when the tax system incorporated subsidies for a select few and correlative penalties for the rest of us.

Many modern (political) debates on tax reform appear to assume that the level of wealth is independent of the distribution of the tax burden. Too often the question of "fairness" in tax legislation is taken to mean that the distribution of tax burdens has no allocative effects. Nothing could be further from the truth. Taxes can create distortions that reduce wealth. The question is how to minimize these distortions. A sound tax, which funds only needed public goods, will always be subject to some evasive response by private parties. But these are minimized by a system that does not seek to reshape basic preferences and that tries to insure that every public expenditure is worth more to every party taxed than the revenues that are lost. Tax neutrality represents one constraint on government policy that is designed to keep government activity as close as possible to the Pareto ideal.

B. Costs of administration. Any tax system costs money to administer, costs which are incurred both by the state and by private parties. These costs are incurred in order to bring about a simple transfer of wealth between two sides. The smaller their sum, the greater the fraction of the social product that is left for use in either public or private hands. In general, simple rules are far easier to administer when mass transactions are at issue. A tax system that demands extensive information to implement is inferior to one that has greater ease of operation by depending upon reliable and measurable facts on the public record. If the costs of collection are high, it increases the likelihood that the taxation (and its associated expenditure) will not meet the requirements of a proper Pareto-superior forced exchange.

C. Rent-seeking. Taxation should be made as invulnerable to political pressures and manipulation as possible. The dangers of faction and abuse of power tend to unfairly skew the terms of forced exchanges. Again, there is nothing unique about taxation. The general theory of property rights makes it clear that the opportunity for political, rent-seeking behavior is greatest where property rights are ill-defined. Taxation can easily be fitted into the general model of property rights, for any tax creates a *lien* for the benefit of the state upon the assets of all its citizens. Talk of a "tax lien" is no loose or florid analogy. Taxes are "floating liens" (parallel to those obtained by many private lenders) that the state may quickly convert into a fixed lien against particular property when taxes are unpaid. That lien can then be foreclosed and the property sold to satisfy the underlying obligation.[12]

The characteristics of the floating lien influence the operation of the system. The greater the freedom state officials have in specifying the contours of the floating lien, the greater the resource losses the tax system will create. Again, the theory of property rights accounts for this result. Ill-defined rights confer discretion upon public officials. Discretion in turn means that there is greater uncertainty in the income stream of all individuals taxed, which typically implies a welfare loss since most persons are risk averse. Discretion also creates a common-pool problem not dissimilar to that which applies to fisheries or oil and gas. It is one thing for A, B, C, and D each to owe $25; it is quite another thing for A, B, C, and D to owe jointly $100. In the first situation (that of separate debts in equal amounts) the obligations are well-defined, and the persons involved cannot shift by political action the burden of their obligations to others. In the second situation these obligations are ill-defined, and each person is tempted to spend real resources (up to $25) to deflect some or all of the common debt to others. Those strategic expenditures in the aggregate produce dead weight social losses, which threaten the integrity of the taxation bargain: when taxation shrinks the social pie it cannot be a Pareto-superior forced exchange, no matter what the final distribution of rights.

IV. IMPLEMENTING THE BARGAIN

It now remains to consider the extent to which both substantive and institutional rules can implement these primary features of the taxation bargain. I believe that they can. In some cases, the theory points to very clear results, while in others the question of trade-offs (chiefly between adminis-

[12] Hence the demand for the pro rata distribution of the surplus. See Coleman, *The Economic Analysis of Law.*

trative costs and incentive effects) is far more vexatious. What follows is an effort to apply the basic theory to some of the central features of a sound system of taxation.

A. Cash or kind? The first proposition is that taxes should normally be collected in cash, not kind. With virtuous governments and virtuous individuals, the choice in the form of taxation might seem unimportant. The right assets will move from private to public hands but only in ways that surmount the standard public goods problems. But as ever, the choice of institutional form depends critically upon the ability to resist the full panoply abuses, both public and private, that arise when information is imperfect, the costs of monitoring are positive, conflicts of interest are rampant, and the temptation to siphon off public resources for private benefit is pervasive.

In this imperfect world, the monetary tax has major operational advantages. From the side of the individual taxpayer, the imposition of a strict monetary obligation allows him to decide which resources to sell or mortgage in order to pay the tax. In general we can expect that the individual will preserve from the taxing authorities those assets to which he attaches the highest subjective valuation, that is, those in which his value in use is greater than the market value.[13] As all persons are in the same position with respect to the tax obligation, each person is able to appropriate the entire surplus value to himself, which in turn reduces the likelihood that the wrong resources will be taken for public use. The upshot is that it is more likely that the operation of the system will be Pareto-superior, with a pro rata division of the surplus, than will a system that allows the government to select the assets to be taken from each individual.

The constraint that taxes be paid in money also imposes an important limitation upon the discretion of government officials. If the government were allowed to impose taxes in kind (say 25 percent of the goods owned by the taxpayer), individual taxpayers would have powerful incentives to influence the choice of assets taken by the government. Government officials might (improperly?) threaten to take assets with high subjective values in order to increase the opportunities for obtaining bribes.[14] Thus, if one's home is worth $25,000 in the open market, but $35,000 subjectively to the taxpayer, the ability to take the home instead of cash creates the opportunity for bribes between one dollar and $9,999. Side payments in that amount

[13] On the importance of subjective value generally, see Timothy Muris, *Cost of Completion or Diminution in Market Value: The Relevance of Subjective Value*, 12 J. Legal Stud. 379 (1984).

The question of subjective value plays an important role in the law of eminent domain, on which I have written generally; Richard A. Epstein, Takings: Private Property and the Power of Eminent Domain (1985).

[14] On this process generally, see Benson & Baden, *The Political Economy of Governmental Corruption: The Logic of Underground Government*, 14 J. Legal Stud. 391 (1985).

would leave both the official and the private party better off, but would result in large systemic social losses. These opportunities for game playing are not restricted to the immediate relationship of taxpayer and public official, for it is possible to envision deals in which bribes are paid to induce tax collectors to levy against the key assets of a business rival or a political opponent.

The possibilities for abuse do not stop when the specific assets are taken into government hands. The management and sale of unwanted assets must proceed on a gigantic scale, and offer ample opportunities for negligence, corruption, favoritism, and abuse. Even if these forces could be banished from public life, the administrative costs of the system would still be very high. A simple rule that all taxes be paid in cash provides a powerful and workable constraint upon both public and private misbehavior.

It is easy to overlook the importance of these institutional arrangements in dealing with matters of taxation. Robert Nozick has argued, for example, that "Taxation of earnings from labor is on a par with forced labor."[15] His conclusion is true if the only relevant political inquiry is whether or not taxation involves the use of coercion. A pure libertarian could therefore condemn taxation and forced labor with equal vigor.[16] Yet once *any* forced exchanges are admitted into the system, then Nozick has only answered the first question: Is there a taking of private property? He does not begin to answer the question of whether some greater benefit is provided in return. On that question the differences between taxation and forced labor, which are widely understood intuitively, become critical. The problems of private evasion and public abuse just mentioned make it clear that a system of ordinary taxation is more apt to generate Pareto-superior taxes than a system of forced labor. The differences here are as substantial as those between a system of conscription (a tax in kind on some individuals) on the one hand and a volunteer military force paid for by general tax revenues. Nozick's categorical condemnation of all coercion provides no framework for ranking alternative systems of taxation or, indeed, any government action, in any responsible way.

There are, of course, some limits to the principle that the government's sole source of resources should be monetary taxation. While taking labor is generally unwise, taking property may not be. Simple taxation may not work, for example, when holdout problems threaten to block the acquisition of specific land for a public highway. Occasionally, the eminent domain power to take property must be invoked. But as with taxation, the object of the law is to implement Pareto-superior forced exchanges with a pro rata division of

[15] Robert Nozick, Anarchy, State and Utopia 169 (1974).

[16] Nozick has some sense of the point for he notes that he equivocates by the use of the phrase "on a par with," which may mean simply "is" or is "a kind of"; *id.* at 169, note*. Yet he never explores any of the functional reasons that make for the differences.

the surplus. In principle, the eminent domain power should strive to work toward a distribution of benefits and losses identical to that achieved by a well-functioning tax. The compensation requirement operates to insure a return of benefits to the individual equal to the value of the property surrendered, which are then augmented by his pro rata share in the social surplus (if any) generated by the government activity. Eminent domain is far more expensive to operate than ordinary taxation. The in-kind nature of the appropriation calls forth an elaborate network of protections for private parties, not the least of which is the right to a judicial determination of the value of the property taken. Yet here, litigation is so expensive and valuation so controversial that much abuse remains.[17] A well run society subordinates the use of eminent domain to the use of taxation.

B. The Comprehensive Tax Base. The Lockean theory offers guidance on the appropriate base for taxation. The ends of government are to protect a person in all his "life, liberty and estates," so that presumptively everything of value protected by government is subject to taxation. That intuition is captured in the two standard definitions of income, those of Haig and of Simons, which today generally guide theoretical discussions of taxation. Haig's definition of personal income was "the money value of the net accretion to one's economic power between two points in time."[18] Simons' definition made more explicit the place of consumption in the general picture: "Personal income may be defined as the algebraic sum of (1) market value of rights exercised in consumption and (2) the change in value of the store of property rights between the beginning and end of the period in question."[19] Both these comprehensive definitions are required by the Lockean theory, for both advance tax neutrality by preventing distortions in the prior patterns of individual preferences.[20] To be sure, the definition does not answer every question about the proper base for taxation. There is, for example, a debate over whether casualty losses for nonbusiness property

[17] One obvious point is that the government normally need not compensate the owner anything for the legal and business costs incurred when condemnation takes place. See, e.g., Bodcaw v. United States, 440 U.S. 202 (1979). For the complications in setting the right rule, see Ayer, *Allocating the Costs of Determining "Just Compensation,"* 21 Stan. L. Rev. 693 (1969), and Epstein, Takings: Private Property and the Power of Eminent Domain, Ch. 4 (1985).

[18] Haig, The Federal Income Tax 7 (R. Haig ed. 1921).

[19] H. Simons, Personal Income Taxation 50 (1938). Professor Bittker has observed that Simons regarded Haig's definition "as interchangeable with his own." See B. Bittker, *A "Comprehensive Tax Base" As a Goal of Income Tax Reform*, 80 Harv. L. Rev. 925–932 (1967). It is not clear why Simons chose to measure consumption by market value instead of subjective value. For a discussion of the difficulties with subjective variations, see *infra* at section (c).

[20] Simons was no stranger to political theory; see his *Some Reflections on Syndicalism*, 52 J. Pol. Econ. 1 (1944), reprinted in H. Simons, Economic Policy for a Free Society (A. Director ed. 1947).

should provide a deduction to their owner.[21] While it might be thought that the casualty deduction for personal property compromises the rate base, the opposite conclusion seems correct.[22] The destruction of property is a net decrease in wealth, which cannot be regarded as consumption. Hence, by the Haig–Simons definition it reduces income and thus functions as a deduction against income from other sources. Similar but more difficult questions arise with, for example, medical deductions and charitable contributions.[23] But these definitional problems can be given plausible answers, and in any event cannot be allowed to obscure the important agreement over many different kinds of income – wages, interest, royalties, rents, dividends, capital appreciation – now reached by the tax system,[24] and which under any view compromise the huge bulk of any tax revenues.

Finding the appropriate definition of income can be regarded either as an end in itself or as the start of a more complex inquiry into tax policy. In theory, it should clearly be the latter. Once the comprehensive definition of gross income is established, the issue is whether there is any deviation from the original tax base that satisfies dual constraints of Pareto superiority and pro rata division of the gain. A moment's reflection should indicate that further refinement is not only possible, but strictly necessary.[25]

The most vivid illustration is imputed income, that is, income which comes from the productive use of one's talent or property without the interposition of any market transaction, such as a sale or a salary. The consumption of home grown food and the value of services about the home are two common illustrations. In principle, all imputed income falls within the taxation base, as the principle of tax neutrality opposes any differential tax cost on market income that will bias individuals toward nonmarket uses

[21] Internal Revenue Code (I.R.C.), § 165(c)(3).

[22] Epstein, *The Consumption and Loss of Personal Property under the Internal Revenue Code*, 23 Stan. L. Rev. 454 (1971).

[23] See William D. Andrews, *Personal Deductions in an Ideal Income Tax*, 86 Harv. L. Rev. 309 (1972).

[24] See I.R.C. § 61(a), which contains a general reference to gross income and a long list of included items. Cases of the fringes include such matters as the treatment of punitive damages, on which see Glenshaw Glass Co. v. Commissioner, 348 U.S. 426 (1955), covering exemplary damages in fraud cases and the punitive portion of recovery in an antitrust action. The doubts about recovery stemmed from the definition of income used previously by the court in Eisner v. Macomber, 252 U.S. 189 (1920): "'Income may be defined as the gain derived from capital, from labor, or from both combined,' provided it be understood to include profit gained through a sale or conversion of capital assets." Note, this definition excludes the gain received in *Glenshaw Glass* which is, however, taxable under the Haig–Simons definition.

[25] Once the point is grasped, it becomes clear that any attack on the Haig–Simons approach that is content to point out deviations from the system is not sufficient to condemn its usefulness. For an illustration of the error, see Bittker, *A "Comprehensive Tax Base" As A Goal Of Income Tax Reform*, 80 Harv. L. Rev. 925 (1967), effectively criticized in Musgrave, *In Defense of an Income Concept*, 81 Harv. L. Rev. 44 (1967), which, while it does not adopt the explicit framework here, takes much the same approach.

of their own labor. But tax neutrality has to yield something to administrative necessity, for it is simply impossible to value every pleasure or benefit that people routinely enjoy. The intangibles are overwhelming. Do persons with happy home lives get taxed at a higher level than persons who come from broken homes? Should government officials stand over the shoulders of all citizens in order to tax the value of home repairs? The opera fanatic who buys a box seat for $50 may in fact be willing to pay $75, but should $25 in consumer surplus be subject to taxes? Should the worker who loves his job be taxed more heavily than his companion who does not? Should the person who enjoys watching the sunrise or quiet contemplation be taxed for that as well?

It takes little theoretical insight to realize that the problems of monitoring and evaluation doom from the start the enterprise of taxing all imputed income. The conclusion follows a fortiori given that sound rules of taxation are designed to curb both individual and government abuse. Excluding imputed income is a flat violation of tax neutrality in that it biases the tradeoff between market and nonmarket transactions, and between work and leisure. The critical question is whether that violation can be justified. The relevant inquiries are two: first, will the total size of the pie be greater and, second, will the relative *individual* tax burdens remain unaltered?

The answer to the first question seems to be "yes." The administrative savings from excluding imputed income from the tax base are enormous. The need to value nonmarket goods is eliminated, while the discretion vested in the taxing authorities and the size of the government apparatus are both reduced. Revenue neutrality can be preserved, even if tax neutrality is not, by raising the rates on taxable income somewhat to offset the shortfall that might otherwise arise.

The answer to the second question is also "yes." At first blush, the exclusion of imputed income looks mischievous if one assumes, as does Nozick in *Anarchy, State and Utopia*, that one person enjoys going to the movies and the other watching the sunset, so that the first pays taxes on the income earned to pay for the movie ticket, while the other pays no taxes to watch the sunset. Yet the possibility of having the person with the greater total income paying the smaller taxes becomes far less likely given, as seems the case, that everyone has a large market basket of imputed income transactions that are kept out of the tax system. It matters little that the contents of each basket will differ so long as their total value is about the same. Every person has large amounts of imputed income, some from a good job and some from a good homelife. There is no group of individuals that has a monopoly on imputed income, as most people consume *both* some movies and some unpurchased leisure.

Excluding imputed income does not appear, therefore, to create any systematic bias across groups of individuals in the society at large, and is not obviously correlated with race, sex, age, religion, occupation, or even income levels. The rank ordering of individual utility tends to be preserved after the tax base is redefined. Those persons who would pay more in taxes with the comprehensive base continue to pay more even after imputed income is excluded. This diffusion of benefits and burdens also has important political ramifications, for it is hard to see the interest group that would make the inclusion or exclusion of imputed income the target of rent-seeking or factional politics. It is too hard for the players to identify those persons who would win and lose from the shift.

The recognition that imputed income falls outside the tax system thus appears to produce overall gains evenly shared by the population at large, at least with the approximate guesswork that must be made on matters of this scale. The loss of tax neutrality is a serious matter, but the best solution is simply not obtainable. The revenue needs of the state can be met by a higher tax on the smaller base, with no obvious redistributive consequences. To be sure, some items of imputed wealth are of sufficient magnitude that they might be brought back into the rate base – the rental values of private homes is an obvious candidate for the wrinkle upon the wrinkle. Yet suppose the trade-off between administrative costs and tax neutrality here is incorrectly drawn under the current system. Still the question seems close, so that the costs of error are relatively low. A society can easily flourish with such unresolved ambiguities.

C. When to tax: the realization and recognition of income. The second set of qualifications of the rate base have to do with the nature of the receipts and expenditures that trigger collection of the tax. As Simons himself repeatedly stressed, the proper definition of income does not depend upon any prior "receipt" or realization of gain. To illustrate his proposition, if I own 100 shares of IBM stock purchased at $100, but which are now worth $125, then $2,500 in gain is income whether I choose to hold or to sell the stock. Ignoring the transaction cost, holding the stock is no different from a sale and a repurchase of the same securities. The comprehensive definition of income thus includes $2,500 in gain (and would recognize a like amount in losses) whether or not the stock has been sold, and thus minimizes the ability of the strategic maneuvers of private parties to undermine the system.

Yet here, too, the concern with tax neutrality has generally not been strong enough to carry the day in all cases. Every modern tax system works off the concept of receipt of property that is immaterial under the Haig–Simons definition of income. Receipt of wages or the realization of gain, as from the sale of stock, has been generally regarded as a condition for the recognition

of gain and for the levying of a tax. The effect in all instances is to delay the *time* for paying the tax until after the increment to wealth or consumption is obtained.

The justifications for the delay are practical in origin but theoretical in their significance. Taxation of unrealized appreciation, for example, can pose real questions of valuation given the absence of market transactions. These are themselves quite insuperable when dealing with real estate or with shares of a closed corporation, when there is no ready market. They are of some modest concern with the shares of public corporations. In both cases, there is also a nagging question of how some taxpayers will raise the cash to pay the tax, for it is quite possible to be cash poor and capital rich. Imposing the tax before realization of the gain can force a taxpayer to go to the market, to mortgage the shares or to sell them at an inopportune time. These transactions can be costly and inconvenient and could disrupt business objectives.

All this is not to say that there are no disadvantages in delaying taxation until the sale of shares or the receipt of cash. Where certain assets appreciate substantially, the clever taxpayer may be able to obtain cash today, without payment of the tax, by the use of a nonrecourse loan, secured by the shares of stock.[26] The system of delayed taxation also tends to discriminate against earned income and in favor of investment income. And it leads to genuine complications of income in periods of inflation or deflation as well. Nonetheless, the administrative burdens of the Haig–Simons definition are too great in a system that must collect taxes from millions of individuals. The deferment of taxation to a later period probably works a smaller shift in wealth between persons than it does between different kinds of income, and it spares everyone the enormous administrative costs that would otherwise be incurred. On balance, deferment looks wealth-enhancing and, hence, Pareto-superior, especially in light of the next feasible alternative. One might want to dispute that point,[27] especially as regards shares of publicly held corporations; but the debate here would not be on the soundness of the

[26] See Woodsam's Associate v. Commissioner, 198 F. 2d 357 (2d Cir. 1952). See, generally, Marvin Chirelstein, Federal Income Taxation §13.02 (1985).

[27] See, e.g., Slawson, *Taxing Ordinary Income the Appreciation of Publicly Held Stock*, 76 Yale L.J. 623 (1967). The opposite side of the coin is the immediate recognition of unrealized losses. Yet the parallel is not precise because even with the realization requirement a taxpayer could always choose to sell in order to recognize the loss. Note too that this strategic power has in more recent times been used by investors in various kinds of options. The basic principle is to invest in a portfolio of options to buy and sell at different prices and dates. As the portfolio is negatively correlated, some items will go up and others down. The losses can then be realized, while the gains are postponed. A forced accounting without realization at the end of the tax year defeats the strategic maneuvers of taxpayers. The problem does not exist in the same degree for stocks because their prices tend to be positively correlated and they have lower volatility.

general criteria, but on their application to a particular case – a second order problem.

The question of realization is at bottom a question of timing: *when* should certain forms of wealth be taxed. Often it is not sufficient to defer taxation until the receipt of wealth. Stated in standard tax language, the realization of gain or loss need not necessarily be the occasion for the "recognition" of the gain. Taxation can be deferred until some later point in the wealth production cycle, usually when the property received is reduced to cash or its equivalent. For example, a shareholder contributes property to his wholly owned corporation in exchange for an issue of its shares; the form of wealth has changed, and gain has been realized from the sale of the thing. But in general, the recognition of the gain for tax purposes will await the sale of the shares for cash by the shareholder, or the sale of the thing by the corporation.[28] Or two pieces of real estate are exchanged and both parties realize a profit, in that the property received is worth more to each party than the cost of the property surrendered. But, again, the deferred recognition of the gain seems appropriate.[29] The practical problems of taxation are evident, and there is a risk that the tax cost might well prevent what would otherwise be a beneficial economic exchange. The gains can still be taxed, either in the form of rental income or upon a subsequent cash sale.

The treatment of unrealized losses raises the same issues in reverse. Normally depreciation of property used in a trade or business precedes its sale; the refusal to make any allowance for depreciation before recognition could result in the bunching of losses in the year property is sold or abandoned. Yet taking these losses into account long after they are sustained produces serious economic distortions. In principle, the proper response is to value real estate at the beginning and end of the tax period, so that losses in depreciation can be combined with appreciation (or depreciation) in property value attributable to market forces to reach a net figure of taxable income or loss. This valuation process is cumbersome and unreliable. One could ignore depreciation over time, as one ignores appreciation (or depreciation) for changes in market value of stocks. But there is a difference: market forces can move value in either direction, while age and obsolescence exert a systematic downward pressure.

It seems, therefore, defensible (and probably correct) to use some simple taxation formula to capture that systematic downward shift in value in order to reduce the anticipated gap between economic and tax definitions of income, all at a tolerable administrative cost. Today, the tax system itself offers a bewildering array of options for depreciation not required by the

[28] See I.R.C. § 351.
[29] See I.R.C. § 1031.

basic theory.[30] The needless discretion they vest in taxpayers should be eliminated because of the strategic game playing (and consequent social losses) it invites. But some simple formula for depreciation over broad categories of property is surely appropriate,[31] even if the present menu of options is not.

The same logic applies to income received today for services to be performed tomorrow. Consider the receipt of the right to future payment from long-term service contracts. The associate who is made a partner at a law firm, the professor who receives tenure at his university, the federal judge appointed for life might be said to have received fair market value of the future income stream that attaches to his employment. Certainly there are chances that he will quit, be fired for cause, or die, but nonetheless the receipt of the contract right from the employer is not a matter of indifference to the taxpayer. There is a clear receipt of a thing of value (the contract right), and the net change in wealth is substantial, even if there can be genuine quarrels on the matter of valuation. But should the income stream, however valued or discounted, be taxed in the year the right is acquired? Here the problems of getting the money to pay taxes on a future and uncertain income stream are acute, for what bank will lend on the strength of the bare contract right that the government is so eager to tax. Nor is it possible at the moment that job rights vest to make an accurate assessment of the future income projection. Partnerships have good and bad years; universities have pay freezes; judges' salaries are tied up in politics. The receipt of the long-term contract right is a poor time to impose the tax. The reduction of the right to the cash is the appropriate instance, as the law now generally provides. Taxation of in-kind benefits (like the collection of taxes in-kind) is perilous and unnecessary business. A little waiting reduces uncertainty and administrative costs. The shift seems to meet the dual requirements of Pareto superiority and pro rata distribution of the gain.

D. Income or consumption taxes. The permutations for the recognition and deferment of gain can be extended one step further, since many people argue that the receipt of income should not be an occasion for taxation; instead, taxation should be further postponed until the moment of consumption. The shift to a consumption tax may appear revolutionary, but a few key structural changes in the tax laws could bring much of it about in a modest way. One change would be to forgive the gain on securities when the

[30] See I.R.C. §§ 167, 168.

[31] In a world without inflation, the proper formula for depreciation would reduce the value of a long term asset (which has uniform value in use) by the present discounted value of the last remaining year's use, not by straight line depreciation, which simply treats the value of each year's use equally. Again, the error, while large in principle, may not shift wealth between individuals, even if it could (by overencouraging investment) result in some allocative losses. See Marvin Chirelstein, Federal Income Taxation ¶ 6.07 (4th ed. 1985).

proceeds of sale are reinvested in other forms of investment property. Another would be to allow a deduction for savings. These changes would ignore the proper tax treatment of consumer durables,[32] and thus are not complete. Once the choice between the income and consumption tax is seen as one of the timing of gain, there is no obvious need to choose one pure type over another. An intermediate position may prove to be socially preferable. The question of whether the consumption tax is superior to the income tax is vexed indeed,[33] but here, as before, the structure of the inquiry is the same as in all other instances: is the change Pareto-superior with an anticipated pro rata division in the surplus? The proration requirement seems easily satisfied, so that the answer must rest on a detailed examination of the administrative and incentive effects of the two systems and the intermediate position, a task which is largely beyond the scope of this paper.

E. Pathological cases: tax subsidies. Many other deviations from the Haig–Simons definition are not consistent with the dual constraints of the general theory. The Lockean theory places a flat prohibition upon implicit transfers from the state that leave one group better off and another worse off. That prohibition reaches tax subsidies like all other transfers. Any argument that taxation is an efficient mode of implementing subsidies therefore misses the mark, because it presupposes that subsidies themselves are legitimate, a clear violation of the basic political norm.

Within this framework, a large number of special provisions of the Internal Revenue Code are simply improper. The rapid depreciation write-offs for low income housing represent one obvious subsidy that is indefensible on these grounds. The grant of percentage depletion allowances on natural resources in excess of original cost is also indefensible. The decision to keep municipal bond income outside the tax base is similarly suspect. The willingness to allow shareholders who make charitable gifts to deduct the full value of appreciated property without paying any tax on the interim gain is likewise inappropriate.[34] The free stepped-up basis at

[32] See Graetz, *Implementing a Progressive Consumption Tax*, 92 Harv. L. Rev. 1575, 1613–23 (1979). Note, there would be a real disincentive to purchase a home if the full purchase price was regarded as consumption in the year of acquisition. It is possible to defer the collection of the tax over a number of years to reflect the deferred receipt of the benefit. This delayed consumption tax on durables is best understood as the mirror image of the deferred deduction for depreciation now allowed under the income tax.

[33] See, e.g., Andrews, *A Consumption-Type or Cash Flow Personal Income Tax*, 87 Harv. L. Rev. 1113 (1974); Graetz, *Implementing a Progressive Consumption Tax*, 92 Harv. L. Rev. 1575 (1979); Kelman, *Time Preference and Tax Equity*, 35 Stan. L. Rev. 649 (1983). Note, throughout I have ignored the flat value-added tax, even though it is surely a plausible substitute for either income or consumption taxes.

[34] See I.R.C. § 170(e). In its simplest form, the taxpayer contributes shares that cost $100 but which are now worth $1000 to charity. The right calculation imputes the gain of $900 and then allows the deduction of the full $1000. The deduction appears to be $100, but is in reality more if the imputed gain is taxed at capital gains rates and the deduction is permitted against ordinary income.

the time of death is also a source of reproach.[35] Many of these gains are available only to a tiny segment of the population; many violate the principle of tax neutrality while creating additional administrative costs. The existence of a number of close cases should not blind us to the power of the theory in disposing of many easy cases, with enormous importance to the system as a whole.

F. Rate structures. The general theory has important implications for the rates of taxation. In traditional tax parlance, these rates can assume one of three forms. They can be regressive, progressive, or proportionate. All three systems share the basic belief that the amount of taxes owed should increase with the level of income earned. (I ignore here capitation taxes that call for all persons to pay a fixed amount of taxes, regardless of income.) All three systems assume that the rate base can be defined to exclude items of imputed income, and to respond to the demands for the deferred recognition of taxable income. They differ in the way they relate the increase of taxes to the increase in income. Regressive taxes increase more slowly than taxable income; proportionate taxes increase at the same rate; and progressive taxes, today the dominant form, increase more rapidly. From the Lockean perspective, a strong case can be made that the proportionate or flat tax is the only acceptable version, as Locke hints in his own brief account of taxation.

The first point in the argument is to note that within the Lockean world, the redistribution of income through the tax system is an unacceptable function of government, for no exchange can be Pareto-superior if it leaves some persons worse off.[36] Likewise, the requirement of the pro rata division

[35] See I.R.C. § 1014. To explain: if property was bought at $100 and is now worth $1000, the $900 in gain escapes taxation at death, creating the obvious incentive to postpone realization. One solution might be to allow the basis to carry over through death, so that the gain can be obtained on sale. That proves in essence to be quite unworkable if the property is held in trust, because there is no sane formula for the allocation of basis among the beneficial interests. The immediate income tax at death, coupled with the abolition of the estate tax, is the best solution.

[36] Here the learned can debate Locke's own position on the matter. The Second Treatise, his most influential work, contains no hint of any obligation to redistribute wealth that is enforceable by the state. In his criticism of Filmer, he has written passages that indicate the power of charitable obligations to the needy. Yet it is quite clear that even here he regards charity as a different domain from justice, and does not take into account the possibility that imperfect obligations of conscience are different both from legal obligations and pure consumption choices. The relevant passage is Chapter 4, ¶42, which reads as follows:

> 42. But we know God hath not left one man so to the mercy of another that he may starve him if he please. God, the Lord and Father of all, has given no one of His children such a property in his peculiar portion of the things of this world but that he has given his needy brother a right in the surplusage of his goods, so that it cannot justly be denied him when his pressing wants call for it; and, therefore, no man could ever have a just power over the life of another by right of property in land or possessions, since it would always be a sin in any man of estate to let his brother perish for want of affording him relief out of his plenty; for as justice gives every man a title to the product of his honest industry and the fair acquisitions of his ancestors descended to him, so 'charity' gives

of the gain entails that the just tax should not alter the rank ordering by wealth of individuals within the society. All of this is said with the understanding that *if* redistribution were an appropriate social goal, then taxation would be superior to ad hoc forms of direct regulation in bringing it about.[37]

If the questions of income redistribution are put aside, the concerns identified above are surely dominant. The progressive tax scores badly on the ground of tax incentives. One obvious point is that the higher marginal rates encourage individuals to conceal their revenue from the government, thereby increasing the portion of the tax burden that falls on others. Yet even if tax compliance were perfect, distortions would remain, for able high-income individuals will reduce the level of services they offer for sale because of the relatively low after tax return.

The progressive tax also scores very poorly on administration. The question is not one of computation. Instead, it involves the elaborate network of collateral rules that must be developed in order to protect the integrity of the progressive rate structure. Income splitting is of no consequence with the flat tax, for the total amount payable to the government is by definition independent of the number of persons to whom it is paid. But it becomes critical whenever any degree of progressivity is introduced into the system: then, to spread the wealth is to minimize the tax burden. Consider the rules that scrutinize the payment of salary to children, that regulate the taxation of income from family partnerships and multiple trusts and corporations, and that regulate the sale of appreciated property by donees. These, and many more, are all necessary government efforts to counter efforts by taxpayers to avoid taxes. Both private efforts to deflect the progressive tax and government efforts to maintain it are dead weight social losses that can be largely eliminated by the flat tax.

> every man a title to so much of another's plenty as will keep him from extreme want, where he has no means to subsist otherwise. And a man can no more justly make use of another's necessity to force him to become his vassal, by withholding that relief God requires him to afford to the wants of his brother, than he that has more strength can seize upon a weaker, master him to his obedience, and, with a dagger at his throat, offer him death or slavery.

It is far from clear how extensive a system of support for the poor this passage calls for. The duty to supply welfare is limited to cases of "extreme want" and not mere poverty. Locke, moreover, nowhere addresses the question of whether the money received has to be repaid should that become possible, and he does not ask whether "justice" and "charity" require the same modes of enforcement. It is quite possible that claims of justice are enforced through the legal system, and those of charity by a network of imperfect social obligations. I have ignored redistribution questions entirely here, but developed the case against redistributivist government policies in Epstein, *The Uncertain Quest for Welfare Rights* [1985] B.Y.U.L. Rev. 201 (1985).

[37] See A. Mitchell Polinsky, An Introduction to Law and Economics, 9–10, 105–113 (1983). The one great unanswered question is whether it is possible to have any effective scheme of income redistribution with a flat tax.

The progressive rate structure also introduces other complications. The constant, unhappy debate over joint returns for married couples arises because the progressive tax collides with the definition of the unit of taxation, be it individual, household, or family. With a flat tax, there is only one table, no matter what the taxable unit. There need be no question of a "marriage tax" when the income of husband and wife are treated as earned by a single person. Income averaging raises similar problems. The bulges in one year are not a reflection of permanent income. The taxpayer who earns $100,000 in each of two years is treated differently from one who earns $200,000 in one year and $0.00 in the other. A flat tax removes the need for further sets of costly provisions. At every point, the progressive structure of the taxes invites thrust and counterthrust. The effort to shift tax burdens is another illustration of the negative-sum games that a flat tax would largely forestall.

Last, there is the problem of rent-seeking through the tax system. Today, the requirement that all individuals in the same tax bracket pay the same tax provides an important brake against factional intrigue. It would be quite intolerable for the government to set individual tax rates for individual taxpayers: general tax brackets check that manifest form of abuse. But progressive taxation still leaves open a degree of freedom that can be exploited by political action. The flat tax gives the government only one degree of freedom: what is the level of the revenues and, hence, of taxation? That discretion it needs. The progressive tax leaves it a second degree of freedom: what is the level of the tilt for which there is no unique answer. It offers the government a discretion that has no social benefit, but the usual social costs. As persons know their anticipated income with some degree of reliability, they can seek to impose the costs of government upon others while keeping the benefits for themselves. Forestalling those struggles (while giving a tolerable match between the benefits and costs of government services) is one virtue of the flat tax. A rule that says you must pay a dollar for the dollar that you wish to exact from your neighbor is not a perfect safeguard against political intrigue, but it acts as an effective constraint. The flat tax has the necessary effect of reducing the total level of expenditures that are funneled through the government and, in consequence, reduces the tax gap between imputed and taxable income. It is an indispensable part of the Lockean program of taxation.

V. THE CONSTITUTIONAL DIMENSION

The last part of the inquiry is what set of institutions are needed in order to stabilize the power of the government to tax. Initially the question is, do we need any kind of constitutional constraint at all? The answer to that question depends (at least in part) on the dominant political mores of the

time. A system (like England of the nineteenth century) may witness the simultaneous growth of two doctrines: the absolute sovereign power of Parliament and a strong social commitment to the principles of laissez-faire. But the question is whether that will endure or, as has been the case in England, whether the opportunities for gain through political markets will in the end undermine the position of voluntary markets. Here there is reason for pessimism: a political victory for voluntary markets does not preclude a future attack, as the opportunity for gains is still present. Yet a victory for government markets sets up entrenched groups whom it is very hard to root out after the fact. As rent-seeking and faction are the major sources of abuse, there is no obvious presumption that the form and objects of taxation should be left to legislative whim, and some strong reason to believe that in the long run the power to tax will be used to increase the size of government.

As a practical matter, however, it is not enough to identify a major source of abuse. The question is whether it is possible to construct at reasonable costs institutional arrangements that will control it. In this instance, constitutional restraints on the power to tax are the only available tool known to political theory. Whether these are obtainable politically or by judicial action is beyond the scope of this paper. What I do address briefly is the question of whether it is possible or desirable to frame constitutional limitations that address each of the six points raised in the proceeding section. The task is one that is undertaken only briefly, and with great trepidation. It seems clear that our own original constitutional structure did purport to impose explicit limitations upon the power to tax,[38] and contains other general provisions (such as the due process clauses and the eminent domain clause) that could be applied to the legislative power to tax.[39] Yet as noted earlier, where suspect classes and fundamental rights (as defined by the current jurisprudence) are not at issue, these limitations have all eroded with time; so today, over vast spheres of economic activity, the power to tax is plenary.[40]

But what kind of strategy could do better? Here it seems best to trace the possible constitutional responses to each of the six issues raised to see what could be done.

[38] See, e.g., Art. I, § 8, cl. 1: "The Congress shall have Power to lay and collect Taxes, Duties, Imposts and Excises, to pay the Debts and provide for the common Defense and general Welfare of the United States; but all Duties, Imposts and Excises shall be Uniform throughout the United States." The provision appears to limit the proper objects of taxation to the provision of various types of public goods, and was tied to the enumerated powers elsewhere in Art. I. Transfer payments, which consume an ever larger fraction of the government, were not part of the business of the federal government under the original scheme.

[39] I have discussed these issues at length in Richard A. Epstein, Takings: Private Property and the Power of Eminent Domain, Ch. 18 (1985).

[40] See, e.g., A. Magnano Co. v. Hamilton, 292 U.S. 40 (1934); Alaska Fish Salting & By-Products Co. v. Smith, 255 U.S. 44 (1921).

A. Cash or kind. In a fundamental sense, this limitation is already contained in the present constitution which, with its implicit distinction between the taxing and the eminent domain power, recognizes that far greater institutional safeguards are required when specific assets are taken from private individuals. One can quarrel with the present law for its inability to subject the full range of takings (e.g. through regulation) to the appropriate level of scrutiny, but those criticisms do not reach the fundamental decision to raise government resources by taxation.

B. The Comprehensive Tax Base. It seems very hard to define by constitutional means the appropriate base for taxation. In general, the federal statutes have taken the very broad base that the Lockean theory requires. Yet that theory is only the first approximation, given the administrative problems, for example, with imputed income. To draft a provision that speaks both to the original base and the permissible transformations of it seems a very hard task indeed. As the level of abuse from this quarter at least seems quite small, the current position is probably an acceptable one. Certainly the constitutionalization of the tax base would, even if on balance desirable, rate very low on the ladder of pressing constitutional reforms.

C. The realization and recognition of gain. There is no plausible reason to enshrine the timing of income in either direction. Thus, there are occasions in which recognition of gain and loss should proceed its realization, and others in which it should follow it. The key limitation is equal treatment across taxpayers, which is by and large observed in the major provisions concerned with the recognition of income and the allowance of deductions.

D. Income or consumption taxes. The choice between these two categories of taxes is simply a special case of the debate over realization and recognition. There is no reason for a constitutional provision to commit us to one form or the other.

E. Pathological cases. Here the case for some constitutional limitation seems more attractive, but nonetheless not clearly compelling. As noted above, there seems little reason for constitutional provisions that regulate the *timing* of income and deduction, above and apart from one that requires equal treatment across taxpayers. Yet there are many types of tax shelter which work not only to *defer* income from taxation, but to exclude it all together: percentage depletion beyond cost recovery, the exclusion of municipal bond income, and the tax free stepped-up basis at death. In principle, it should be possible to draft a provision that requires that each item of income be recognized at some point during the life of a transaction (i.e., from the time that all relevant players start and end with cash). Identifying a tax loophole is not all that difficult, at least in the easy cases.[41] But it is apt to be a fertile

[41] "While theorists may argue about what constitutes preferential treatment, sophisticated taxpayers have not experienced a similar difficulty." Blum, *The Effects of Special Provisions of the Income Tax on Taxpayer Morale*, 1955 Compendium 251.

source of litigation, and may not be an effective restraint upon the behavior of Congress, especially if there is some legislative power to exclude economic income from the tax base altogether. On balance, I think that the inquiry is at least worth pursuing further.

F. Rate structures. It is here that one sees the greatest opportunity for constitutional reform. At the general level, there is much to be said for a simple provision that says that "all income currently taxable shall be taxed at a uniform, flat rate," with no exclusions of any sort. Yet at the very least, it is problematic whether any system of welfare (or, indeed, any other) transfers could survive under this limitation, as the progressive tax is at least one important source of income redistribution. (How important it is depends at least in part on the way in which expenditures are skewed by government.) In any event, whatever the desirability of this provision as a general matter – and I believe it is great and consistent with the structure of the eminent domain clause – it would be a major revolution in taxation and political economy. Whatever doubts might exist about this provision, they do not carry over to the various special taxes (such as the windfall profits tax) which have begun to proliferate in recent years. Throughout the nineteenth century there was an active constitutional tradition that opposed special taxes for general revenues, and that position should be reaffirmed again today.[42] Special taxes offer greater opportunity for faction and present virtually no opportunity for Pareto-superior exchanges. Any income redistribution can be well done without them via the progressive tax. To enforce a constitutional prohibition against these taxes does not in my judgment require any new constitutional provisions, but only a renewed awareness of the importance of the uniformity and eminent domain provisions of the present constitution and the disproportionate impact tests that are explicit in the uniformity provisions and the "just compensation" requirement of the eminent domain clause.[43]

There is, I think, an important lesson that emerges from the foregoing analysis. In a world that operated without frictions of any sort, no individuals should be allowed to take property under the mantle of public authority. If transaction costs were zero, there would be no case for the state. Nozick would have made his case. But the existence of major frictions of all sorts calls for government. Within this environment, it becomes almost indefensible to argue that private property carries with it the absolute right under all circumstances to exclude others from its use. Yet as forced exchanges are allowed, someone must monitor their terms. As the direct measurements of

[42] See Diamond, *The Death and Transfiguration of Benefit Taxation: Special Assessments in Nineteenth-Century America*, 12 J. Legal Stud. 201 (1983).

[43] "The Fifth Amendment's guarantee that private property shall not be taken for a public use without just compensation was designed to bar Government from forcing some people alone to bear public burdens which, in all fairness and justice, should be borne by the public at large." Armstrong v. United States, 364 U.S. 40, 49 (1960).

benefits and costs are hard to make, a test of equal treatment across taxpayers becomes the next best alternative. In large measure, our current system of taxation, notwithstanding its impersonal and impenetrable character, provides much of the protection that is needed against government abuse. There are some changes, such as those which relate to flat and special taxes, that are important and obtainable. And there is a pressing need to get rid of the conscious subsidies built into the tax system. But most fine points of tax theory (on such matters as realization and recognition) are now construed in a way that is quite consistent with sensible Lockean principles. On these further constitutional control would be inadvisable.

CONCLUSION

In closing, it might be useful to summarize the article. Systems of taxation are needed to supply the public goods that are necessary to protect the lives, liberties, and estates of citizens from aggression by their fellow men. In principle, if one could measure perfectly the benefit that each person obtained from government, then the tax should be levied in proportion to the benefit so provided. Yet insuperable obstacles lay in the path of any system of direct measurement, not the least of which is the massive expansion of government that this scheme requires. To avoid these pitfalls, it is necessary to simplify the tax system to make it workable. Taxes are levied largely upon cash and cash equivalents, usually at the time of their receipt in monetizable form. Rates are kept flat and special exemptions and subsidies are rooted systematically from the tax system. The tax system is designed to minimize the inevitable distortions from the tax free world. These simple rules have some very durable properties, in that they guard against the abuses of faction and reduce the administrative power and discretion that is vested in the state. Controlling taxation is, of course, only one side of the problem. A comprehensive system of social control recognizes the need to impose parallel restrictions on the expenditure side of the budget and on the direct regulation of the economy. In one sense, it is easy to identify the most fundamental problem of all: too many resources today are channeled through government. Sensible constraints on the powers of taxation are not a conclusive remedy to all our ills. But they are a step in the right direction.

Law, University of Chicago

Criminal Law, Tax Evasion, Shams, and Tax Avoidance: Part 1 — Tax Evasion and General Doctrines of Criminal Law

JOHN PREBBLE BA, LLB (Hons) (Auckland), BCL (Oxon), JSD (Cornell), Inner Temple

Barrister, England and Wales and New Zealand and Professor and former Dean of Law, Victoria University of Wellington.

This article discusses the meaning of "evasion" and how it is punished under the Tax Administration Act 1994 and the Crimes Act 1961. It considers the simple case of evasion, where there is an unadorned failure to report assessable income. The article looks at cases where the failure to report income is innocent, then moves to cases where the omission is intentional. Finally looked at are, cases where a failure to report assessable income is associated with a scheme to avoid tax or with a transaction that might have minimised assessable income, or turned revenue into capital, but that did not have those sought-after results.

The article has been published in two parts. Part 2, "Criminal Law Consequences of Categories of Evasion and Avoidance", will appear in the June 1996 issue of the New Zealand Journal of Taxation Law and Policy.

1 An Introduction to Evasion

1.1 Innocent and Fraudulent Evasion

Lord Templeman explained the difference between innocent and fraudulent evasion in *Challenge Corp Ltd v CIR*[1]:

> "Evasion occurs when the Commissioner [of Inland Revenue] is not informed of all the facts relevant to an assessment of tax. Innocent evasion may lead to a reassessment. Fraudulent evasion may lead to a criminal prosecution as well as reassessment."

Lord Templeman's "innocent evasion" sounds like a contradiction in terms, because one tends to think of "tax evasion" as necessarily entailing fraud. Indeed, that is the sense in which the term is used in New Zealand's income tax legislation, as explained later in this article. However, the expression "innocent evasion" is a useful term, because it gives a label to a concept that otherwise has no name.

Suppose a waiter pockets a tip, planning to add the sum to the computation of assessable income in his tax return. Some days later, the waiter sends his trousers for

1 [1986] 2 NZLR 513, 561; (1986) 10 TRNZ 161 (PC).

cleaning, forgetting to remove the tip and, indeed, forgetting all about the money. The trousers come back with the pockets empty and the waiter never remembers the tip. The tip is certainly assessable and the waiter is vulnerable to administrative penalties and to interest for under-reporting his income. But he is not guilty of fraud.

In 1993, Judge Willy, a Taxation Review Authority, was faced with a case that involved similar issues, though quite different facts.[2] The taxpayer company filed goods and services tax returns that understated its liability. The returns were inaccurate because, unknown to the company's officers and managers, the bookkeeper employed by the company was incompetent and the company's computer software was defective. As there was no dishonesty the taxpayer was not guilty of fraudulent evasion. In this article, the term "evasion" usually refers to fraudulent evasion.

1.2 Evasion and Avoidance

Evasion, fraudulent and innocent, must be distinguished from avoidance. In a case of evasion, the taxpayer's transactions themselves result in assessable income of a certain amount. The taxpayer declares a smaller income in his or her return, either by innocent mistake (as occurred in Judge Willy's case and in the hypothetical case of the waiter) or as a result of fraud (where, for example, a shopkeeper suppresses certain cash sales).

In a case of avoidance, the taxpayer's transactions, taken by themselves, yield a certain level of assessable income. Suppose that this is the amount of income that the taxpayer includes in his or her return. Tax law starts from the proposition that the revenue authorities have no cause for complaint, the taxpayer's assessable income, as revealed by his or her transactions, has been correctly reported.

But it may be that the transactions should be reconstructed to reflect the substance of the taxpayer's dealings more accurately. This reconstruction results in the discovery of more assessable income or in the disallowance of a deduction. The reconstruction may occur by virtue of a statutory anti-avoidance rule such as s BB9 Income Tax Act 1994 (NZ), or by virtue of Judge-made law such as the United Kingdom ("UK") fiscal nullity rule (or "New Approach" rule as it is sometimes called) that started with *Ramsay v IRC*.[3]

To look at the matter from another angle, anti-avoidance rules apply only to real transactions, that is, to transactions that have occurred and that are not shams.[4] If there is no real transaction, but the taxpayer reports income on the basis that there is a real transaction, the case is one of evasion, not avoidance, albeit that if the under-reporting is by mistake the evasion may be correctly classified as innocent.

Tax avoidance falls into two categories. The first category is avoidance that is effective, in the sense that the transactions in question are not vulnerable to an anti-avoidance rule. In *Challenge Corp Ltd v*

---

2 *TRA Case 118* (1993) 17 TRNZ 804, 15 NZTC 5153.

3 [1981] 1 All ER 865 [1981] STC 174. For a summary of the New Approach, see J Tiley (ed), *Butterworths UK Tax Guide* (11th ed), 1992-1993, pp 10-23 and subsequent editions of the same work.

4 *Richard Walter Pty Ltd v FCT* Butterworths Weekly Tax Bulletin, no 39, 29 August 1995, to be reported in ATR.

CIR,[5] Lord Templeman used the term "tax mitigation" for this concept. The second category is avoidance that is not effective because it is caught by an anti-avoidance rule.

The two categories of avoidance and the distinction between them do not correspond to the two categories of evasion. First, in a case of avoidance the taxpayer's calculation of assessable income will be correct if one starts from the proposition that the taxpayer's transactions are effective for tax purposes (and if one assumes that the taxpayer makes no other mistakes in composing the tax return). If a taxpayer's calculation of assessable income is based on transactions that amount to effective avoidance or to mitigation, then the taxpayer's calculation of assessable income is correct and an assessment should be issued on that amount of income. If the taxpayer's transactions are undone by an anti-avoidance rule, the taxpayer's calculations made on the basis of the transactions that were actually carried out remain correct within that context, but those calculations are replaced for assessment purposes by calculations based on other assumptions.

Where there is evasion, on the other hand, the taxpayer's calculation of assessable income is wrong ex hypothesi, whether the evasion is innocent or fraudulent. There is no need for the law to call in aid an anti-avoidance rule to adjust the taxpayer's assessable income. The income as reported by the taxpayer is wrong anyway, usually because one or more items of income have simply been omitted from the computation.

5 [1986] 2 NZLR 513, 561; (1986) 10 TRNZ 161 (PC).

1.3 Offences under the Tax Administration Act 1994

"Evasion" is not defined in New Zealand's income tax legislation. Despite the absence of a definition, the Tax Administration Act 1994 provides in s 186:

> "If any taxpayer evades, or attempts to evade, or does any act with intent to evade, or makes default in the performance of any duty imposed upon the taxpayer by this Act or the Income Tax Act 1994, or the regulations made under those Acts, with intent to evade, the assessment or payment of any sum which is or may become chargeable against the taxpayer by way of tax (which sum is referred to in sections 187 to 191 as the 'deficient tax'), the taxpayer shall be chargeable, by way of penalty for that offence, with additional tax (referred to in sections 187 to 193 as 'penal tax') not exceeding an amount equal to treble the amount of the deficient tax."

Section 186 recognises four separate offences:

(1) Evasion proper;
(2) Attempted evasion;
(3) Acts done with intent to evade; and
(4) Default in performance of a statutory duty with intent to evade.

For convenience, all of these offences are sometimes called "evasion". Only offence (4) necessarily entails a breach of a specific statutory or regulatory term.

The Taxpayer Compliance, Penalties, and Disputes Resolution Bill 1995 was introduced to the New Zealand House of Representatives on 3 October 1995. When the Bill is enacted, it is planned that provisions relevant to offences will come into force from 1 April 1997. Clause 65 of the Bill repeals Part XII of the Taxation

Administration Act 1994, which includes s 186. Section 186 is to be replaced by a new s 143B. The proposed section reads:

> "A person who evades or attempts to evade the assessment or payment of tax by the person or another person under a tax law commits an offence against this Act."

"Evade" is not defined in the Bill, and therefore bears its ordinary meaning, but s 143B(1) adds a number of precise offences that could, generally, be thought of as examples of evasion. For example, s 143B(1)(c) refers to knowingly providing misleading information to the Commissioner of Inland Revenue ("the Commissioner"). The proposed s 143B(4) changes the penalty for evasion to fines and imprisonment, though a proposed Part IX provides for a range of monetary penalties.

Despite these major amendments, the aspects of the law relating to evasion that are relevant to the purposes of this article are unchanged. Accordingly, the discussion in the article is confined to the legislation that will apply until April 1997.

1.4 Definitions

There have been a number of attempts to define or to describe "evasion" in the sense of fraudulent evasion. "Tax evasion" and "defrauding the revenue" are for most purposes synonymous; definitions of one may serve as definitions of the other. Perhaps the most difficult ingredient of evasion to capture in appropriate language is the element of fraudulent intent, though in the end "fraudulent" appears to mean no more than "dishonest".

In the early case of *Stevens v Abrahams*[6] Hodges J said that an intent to defraud the revenue is an intent "to get out of the revenue something that was already in it, or to prevent something from getting into the revenue which the revenue was entitled to get". In a passage in a non-tax case that has often been cited, Viscount Dilhorne said:

> " 'fraudulently' means 'dishonestly' and 'to defraud' ordinarily means, in my opinion, to deprive a person dishonestly of something which is his or of something to which he is or would or might but for the perpetration of the fraud be entitled." [7]

1.5 Acts before Liability is Incurred

It would appear that people can be guilty of tax evasion by virtue of acts that take place before liability for tax has been incurred,[8] although, presumably, for the offence to be complete it would be necessary for tax liability to have crystallised. If crystallisation never occurs there remains the possibility of the offence of conspiracy to defraud or evade.[9]

1.6 Ingredients of the Offence Contrasted with Evidence

One sometimes sees statements to the effect that evasion or fraud *require evidence*

6 (1902) 27 VLR 753, 757.

7 *Scott v Metropolitan Police Commissioner* [1975] AC 819, 839.

8 *O'Donovan v Vereker* (1987) 76 ALR 97 (Full Fed Ct).

9 Ibid.

of a deliberate (and criminal) attempt to deceive the Inland Revenue Department ("the IRD") for the purpose of paying less tax than that which is the taxpayer's statutory liability, or that tax evasion can be defined as the wilful breaking of the tax laws which can be *proved beyond reasonable doubt.*

Such definitions are helpful, except that one should bear in mind two things. First, as with any crime, the ingredients of fraud or evasion may exist even though the authorities cannot find the evidence that is necessary to prove the case. Evasion can exist in a particular case (just as murder can exist in a particular case) even though the authorities can not prove evasion (or murder) beyond reasonable doubt, and therefore cannot secure a conviction. Questions of evidence and burden of proof are not part of the definition of an offence.

Secondly, what is meant by "wilful breaking of the tax laws" in the second definition? If the expression refers only to a breach of a particular provision of the Income Tax Act 1994 or its regulations, then the definition may be only partial. The definition is relevant to the fourth offence ((4) above) in s 186 Tax Administration Act 1994, but the first to third offences ((1) to (3) above) do not necessarily require default in respect of particular provisions of the Act, though in practice they would ordinarily entail at least the filing of a false return of income.

2 Penalties for Fraudulent Evasion

2.1 Tax Administration Act 1994 and the Crimes Act 1961

Penalties for evasion under the Tax Administration Act 1994 include prosecution for filing a false return (s 199(1)(b)), penal tax (s 186), and the publication of names in the New Zealand *Gazette* (s 223). Under the Taxpayer Compliance, Penalties, and Disputes Resolution Bill 1995 the first two penalties are replaced by various provisions in proposed ss 143A(1) and 143B(1), and the third penalty by the proposed s 146.

In a serious case, the authorities will consider prosecution for theft, fraud, or cognate offences under the Crimes Act 1961. Apt provisions include theft (s 220) and conspiracy to defraud (s 257). However, in most cases where tax evasion has occurred (as opposed to questions of attempt or conspiracy) from the point of view of a prosecutor the most promising provision is likely to be s 229A(b), which reads:

> "Every one is liable to imprisonment for a term not exceeding 7 years who, with intent to defraud,—
> "(a) Takes or obtains any document that is capable of being used to obtain any privilege, benefit, pecuniary advantage, or valuable consideration; or
> "(b) Uses or attempts to use any such document for the purpose of obtaining, for himself or for any other person, any privilege, benefit, pecuniary advantage, or valuable consideration."

2.2 Ingredients of Offences Against s 229A(b) Crimes Act 1961

Tax fraud is preeminently an offence that involves documents. Receipts, statements of account, written contracts, and tax returns are clearly documents that are "capable of being used to obtain any privilege, benefit, pecuniary advantage, or valuable consideration". Within the terms of

s 229A(b), "using" such a document "for the purpose of obtaining, for himself or for any other person, any privilege, benefit, pecuniary advantage, or valuable consideration" could occur simply by lodging the document as, or as part of, one's tax return.

For the offence to be complete it is not necessary for the Commissioner to act on the document. It is required only that the offender should "use" the document with the necessary intent. Thus, apart from the need for a "document" the essential ingredients of the offence are "use" and that the use is "with intent to defraud".

The Court of Appeal has confirmed that it is no defence that the document in question came innocently into the possession of the alleged offender.[10]

2.3 Use

There is unlikely to be much difficulty about the word "use". "Use" is clearly not a technical term and, in its general usage, is a word of wide significance.[11] In the unreported case of *R v Adams*[12] Tompkins J posed the question:

> "whether it must be proved that the accused personally used the document in the sense of creating, handling, or signing it, or whether a person uses a document if he directs or authorises a transaction to be carried out which will require the document to be created and used by another."

His Honour had little difficulty in deciding that in the context of s 229A "use" must be construed not only personally but also attributively, to include use by authorisation or direction. There is no reason to construe "use" narrowly. As his Honour said, if the accused "directs that money be paid from the account of company A to the account of company B he, for the purposes of the section 'uses' the documents necessary to make the payment."[13] That is, it is not necessary for the defendant to have handled the document himself or herself, nor to have specifically directed that the document should be created or used. It is enough if the defendant authorised or directed transactions or plans that entailed the use of the document by others. Tompkins J's approach was confirmed on appeal.[14]

2.4 Intent to defraud

In cases of tax evasion the ingredient of "intent to defraud" in offences charged under s 229A Crimes Act 1961 can pose difficulties. One aspect of this ingredient can be disposed of quickly: for intent to defraud to exist it is not necessary for a victim to incur actual loss. As explained in *Adams on Criminal Law*,[15]

> "It is clear that the phrase [intent to defraud] has at its heart the dishonest causing of loss to another . . . Overseas authority indicates that 'intent to

10 *R v Gunthorp* 9/6/93, CA46/93, p 67.

11 *British Motor Syndicate Ltd v Taylor & Son* [1900] 1 Ch 577, 583.

12 (1992) HC Auckland T240/91, pp 22, 23. This was the major prosecution that emerged from the notorious collapse of the Equitycorp group of companies in the early 1990s.

13 Ibid, p 23.

14 *R v Gunthorp* 9/6/93, CA46/93, p 56.

15 3rd ed, Wellington, Brooker & Friend, 1992, para CA246.16.

defraud' is not limited to the causing of loss. It will be sufficient that the victim's economic or pecuniary interests have been deliberately and dishonestly put at risk, whether actual financial loss actually occurs, or was even intended."

Certainly, filing a return that shows a tax liability less than the correct amount would cause loss to the IRD if the IRD acted on it (although, as explained, an offence under s 229A(b) is complete at the stage of "using" the document, whether or not the "use" succeeds). Therefore, the issue in most cases will be: was the taxpayer's intent dishonest? That is a matter on which lay opinion is often as valuable as legal opinion and is a question for a jury.

Whether an action involves "fraud" or "dishonesty" for purposes of the criminal law is never an easy question. The question is particularly difficult in the context of tax evasion, because an alleged offence may entail tax law and criminal law. This article can do no more than identify and briefly discuss doctrines and issues of criminal law that are pertinent to the issue and offer answers to some of the questions that arise. Questions from the general part of criminal law that are relevant include: subjective and objective tests of dishonesty; mistake of law, and the distinction in that context between civil law and criminal law; colour of right; and reliance on legal advice that is wrong. The succeeding sections of this article consider these matters. In respect of questions of subjectivity and objectivity, prosecutions under the Theft Act 1968 (UK) and judgments in such cases, offer some illumination for New Zealand purposes, although New Zealand Courts have declined to follow the leading English case, as will be seen below.

2.5 Dishonesty

The phrase "with intent to defraud" in s 229A Crimes Act clearly refers to the state of mind of the alleged offender. There is a subjective element involved. Thus, for example, if a blind man mistakenly carries his expired travel pass while he is on a bus it would seem that he is not guilty of an offence under the section. Suppose the blind man carries no pass and does not pay. He knows that there is a law requiring payment in terms that apply to everyone, but he holds the moral belief that blind people should not have to pay to travel on public transport. In the UK, this subjective belief will not save the blind man from conviction.[16]

The nature of the test of dishonesty as an ingredient in a fraud case tried in the UK was well explained by Lord Lane in *R v Ghosh*:[17]

> "It is no defence for a man to say, 'I knew that what I was doing is generally regarded as dishonest; but I do not regard it as dishonest myself. Therefore I am not guilty.' What he is however entitled to say is, 'I did not know that anybody would regard what I was doing as dishonest.' He may not be believed, just as he may not be believed if he sets up a 'claim of right' . . . or asserts that he believed in the truth of a misrepresentation [that he has made] . . . But if he is believed, or raises a real doubt about the matter, the jury cannot be sure that he was dishonest.

16 Compare an example given by Lord Lane CJ, delivering the judgment of the Court of Appeal in *R v Ghosh* [1982] 1 QB 10 3, 1063.

17 Ibid, p 1064.

"In determining whether the prosecution has proved that the defendant was acting dishonestly, a jury must first of all decide whether according to the ordinary standards of reasonable and honest people what was done was dishonest. If it was not dishonest by those standards, that is the end of the matter and the prosecution fails.

"If it was dishonest by those standards, then the jury must consider whether the defendant himself must have realised that what he was doing was by those standards dishonest."

It follows from this explanation that by UK tandards it is no defence to a charge of dishonesty that the defendant's action is of a kind that is common, or even customary, among certain groups, if it was dishonest as viewed by ordinary people and the defendant knew that to be so or, probably, would have known it if he or she had thought about it.

In the context of Lord Lane's words, the concept of "claim of right" may perhaps be thought of as a category of absence of dishonesty. An action that involves taking something valuable contrary to a rule of law may be saved from dishonesty for a number of reasons. In the passage above, Lord Lane refers principally to cases where the accused believes that he or she is morally justified. Another possibility is that the accused has made a mistake about the law. For example, a seller wrongly believes that the law permits him or her to repossess a chattel that the purchaser has not paid for. Where an accused mistakenly believes that his or her actions are justified by law, he or she is said to have a "claim of right", which, at least at common law, was a defence to many offences of dishonesty. In New Zealand the concept is known as "colour of right", which is defined in s 2 Crimes Act 1961:

"in relation to any act, ['colour of right'] means an honest belief that the act is justifiable, although that belief may be based on ignorance or mistake of fact or of any matter of law other than the enactment against which the offence is alleged to have been committed."

This last limitation applies also in respect of the UK concept of claim of right.

2.6 Dishonesty in New Zealand

New Zealand Courts purport to apply a more lenient and a more subjective standard to questions of dishonesty than the standard applied in *R v Ghosh*,[18] but it is not altogether clear to what extent the New Zealand standard would result in different decisions in cases of similar facts. In any event, in *R v Williams*[19] the Court of Appeal declined to follow *R v Ghosh*, and said, quoting from *R v Coombridge*:[20]

"In s 222 Crimes Act 1961 [which relates to theft by fraudulent omission to account for monies due to another] no express mention is made of colour of right and the word 'fraudulently' is used on its own. We think that in order to act fraudulently an accused person must certainly, as the judge pointed out in the present case, act deliberately and with knowledge that he is acting in

[18] Op cit n16.

[19] [1985] 1 NZLR 294, 308.

[20] [1976] 2 NZLR 381, 387.

breach of his legal obligation. But we are of opinion that if an accused person sets up a claim that in all the circumstances he honestly believed that he was justified in departing from his strict obligations, albeit for some purpose of his own, then his defence should be left to the jury for consideration, provided at least that there is evidence on which it would be open to a jury to conclude that in all the circumstances his conduct, although legally wrong, might nevertheless be regarded as honest. In other words, the jury should be told that the accused cannot be convicted unless he has been shown to have acted dishonestly."

The result is that in New Zealand an accused's belief in the morality of his or her acts can amount to a defence to a charge that includes fraud as an ingredient. But there remains the issue of the honesty of the belief of the accused. In the UK, if the accused says, "I know everyone else believes that doing x is wrong, but I believe x is morally right" he or she must be convicted according to the law stated in *R v Ghosh*, even if the jury believes him or her. In New Zealand, following *R v Coombridge* and *R v Williams*, if the jury believes the accused the defence should lead to an acquittal. However, a finding that the accused knows that people in general believe his or her actions to be morally wrong must cast some doubt on the honesty of the accused's own purported belief. In the end, there may be less practical difference between the *Ghosh* and *Coombridge* formulations than appears from the words of the Judges.

Be that as it may, in prosecution under s 229A for tax evasion, it is unlikely that the defence will be "I realised that what I did was against the law, but I believed that my actions were morally right". An example of this might be the view that "the Government wastes taxpayers' money and I can do more good for the country by keeping control of my own funds: therefore my failure to pay tax is morally right". In any event, this article is not primarily concerned with evasion of that kind.

It is more likely that an accused will say, "I concede that s 229A proscribes using documents with intent to defraud, but I had no intent to defraud because I thought that my actions complied with the Income Tax Act". Assuming that the taxpayer's actions do not comply with the Income Tax Act this plea raises the issue of mistake of law.

2.7 Mistake of Law and Colour of Right

A *Coombridge* no-dishonesty defence is akin to a mistake of law defence. In the former, in theory an accused can concede that their acts were contrary to the law, but nevertheless be acquitted on the basis that they thought the acts were morally justified. In the latter defence the accused is claiming an honest belief that the law was other than it is. The latter defence is thus tantamount to a defence of colour of right.

As Richmond P points out in the passage quoted from *Coombridge* (above), where fraud or dishonesty is an express statutory element of the mens rea of an offence it is supererogatory to think in terms of colour of right, because a possible reason for absence of dishonesty is a mistake of law. In fact, it seems unlikely that absence of colour of right can be thought of as a separate, independent, ingredient of an offence against s 229A or other offences under the Crimes Act that do not include reference to

colour of right in their statutory definitions.[21]

How is Richmond J's conclusion or, indeed, the concept of colour of right to be reconciled with the general rule that mistake of law is no excuse?

Section 25 Crimes Act 1961 codifies the common law rule that "the fact that an offender is ignorant of the law is not an excuse for any offence committed by him." There is no doubt that s 25 disallows the defence: "I did not know that it was an offence to defraud someone by using a document." Moreover, even if absence of colour of right were an express ingredient of an offence under s 229A Crimes Act 1961, which it is not, the defence would remain disallowed, because, as may be seen from the definition of "colour of right" quoted (above), the defence offers no protection when the accused's mistake relates to the law under which he or she is charged.

The explanation is that the rule *ignorantia juris haud excusat* applies to mistakes of criminal law, not to mistakes of civil law. Thus in *Maintenance Officer v Stark*[22] Beattie J said that a mistake as to one's duties to pay maintenance for a child under the Domestic Proceedings Act 1968, if a mistake of law, was a mistake of civil law, and therefore a defence to a criminal charge of failure to pay maintenance under s 85 of the same Act. Beattie J may be wrong about whether duties under the Domestic Proceedings Act 1968 should be regarded as a matter of civil law, particularly in the context of a prosecution under another section of the same Act. But cases occur where ignorance as to property law excuse offences against property.[23]

On the other hand, in *Waaka v Police*,[24] a case alleging assault on a police constable in the execution of his duty, Cooke P explained in the Court of Appeal that s 25 Crimes Act 1961 is not confined to ignorance of the law under which one is charged. He said:

> "it can be no defence that the defendant, while aware that the person was a police constable, entertained an incorrect understanding of the law regarding the extent of a constable's powers. Section 25 Crimes Act 1961 expressly enacts that the fact that an offender is ignorant of the law is not an excuse for any offence committed by him." [25]

Does this passage mean that, in New Zealand, ignorance of even the civil law cannot be an excuse except in respect of offences where absence of colour of right is an express statutory element of mens rea? The answer is not certain. On the one hand, despite not qualifying what he said in *Waaka v Police*, it seems unlikely that Cooke P intended to go so far as to say that ignorance of civil law cannot found a defence of colour of right. On the other hand, s 25 Crimes Act 1961 is in clear terms that appear to disallow defences based

---

21 Compare with *R v Jones* (1991) 8 CR (4th) 137 (SCC). See *Adams on Criminal Law* (3rd ed), Wellington, Brooker & Friend, 1992, para CA25.08.

22 [1977] 1 NZLR 78.

23 See, for example, Glanville Williams, *Textbook of Criminal Law* (2nd ed), London, 1983, p 456 and cases cited there.

24 [1987] 1 NZLR 754.

25 Ibid, p 759. See also *Olsen v Grain Sorghum Marketing Board* (1962) 55 Qd R 580 (unavoidable mistake of law is no defence).

on ignorance of law, whether criminal or civil.

The answer is uncertain, but assume for the sake of argument that ignorance of civil law may be a defence to a charge under s 229A Crimes Act 1961. The immediate questions are: what is civil law and does tax law count as civil law? The answer is again uncertain, but the results, if not the express reasoning, of a number of cases support the proposition that public law (meaning law that regulates the relationship between citizen and state, including tax law) does not count as "civil law" for purposes of a civil law exclusion from the rule *ignorantia juris haud excusat*.[26]

For the purposes of the rule *ignorantia juris haud excusat* it appears that reliance on competent legal advice puts an offender in no better position than if they make their own mistake.[27] This question, together with the related question of reliance on advice from officials, has received considerable scholarly attention in recent years.[28] The better view appears to be, at least in the U K and New Zealand, that reliance on wrong advice does not extend the exception to *ignorantia juris haud excusat*, albeit that the extent of that exception is by no means clear.

2.8 Mistake of Law, Dishonesty, and Tax Evasion

In the context of s 229A Crimes Act 1961, mistake of law goes to mens rea if it goes anywhere. Because the mens rea of a s 229A offence involves dishonesty, the analytical framework employed in the last section of this article is under some strain. It appears that mistake as to some kinds of laws may furnish a defence to a s 229A charge, in that the defendant was mistaken, and not dishonest. At the same time, it seems that mistake as to some other kinds of laws may not be a defence. There is a lack of logic here. It is not, of course, relevant whether the defendant knows about s 229A. But where there is an allegation of an offence against s 229A by virtue of the defendant's breach of another law, if the alleged offender is genuinely mistaken as to whether his or her conduct breaches that other law, it is hard to see this as dishonesty. Logic reaches that conclusion whether the other law is described as civil, public, or even criminal.

The authorities offer limited help. However, there is at least one case in the reports where the Court has by necessary implication allowed ignorance even of the law under which the defendant was charged to negative a mens rea element akin to dishonesty. In *Wilson v Inyang*,[29] Inyang set up practice as a "naturopath physician". He was charged with "wilfully and falsely" taking the title "physician", contrary to s 40 Medical Act 1858. Inyang's defence, which the Magistrate believed, was that he

---

26 *Grant v Borg* [1982] 2 All ER 257 (HL) (immigration law); *Barras v Reeves* [1980] 3 All ER 705 (Div Ct) (social security law); *Cambridgeshire and Isle of Ely County Council v Rust* [1972] 2 QB 426 (planning and land use law). But see *Maintenance Officer v Stark* [1977] 1 NZLR 78 per Beattie J (social security law).

27 *Cooper v Simmons* (1862) 7 H & N 707; 158 ER 654.

28 See, for example, A J Ashworth, "Excusable mistake of law" [1974] Crim L Rev 652, 657; Glanville Williams, *Textbook of Criminal Law* (2nd ed), London, 1983, p 455; W J Brookbanks, "Officially induced error as a defence to crime" (1993) 17 Crim LJ 381.

29 [1951] 2 KB 799.

honestly thought that he could use the title "naturopath physician" because that expression would not mislead people to think that he was a physician in the usual sense. The difficulty with this defence was that use of the title alone was the relevant ingredient, not use with intent to mislead. Nevertheless, Inyang was acquitted, and held rightly acquitted by the Divisional Court, on the basis of his honest belief. As Lord Goddard CJ pointed out:

> "a man may honestly believe that which no other man of common sense would believe. If he has acted without any reasonable ground, and has refrained from making any proper inquiry, that is generally very good evidence that he is not acting honestly. But it is only evidence."[30]

In the end, the test is subjective honesty. Where the belief of the defendant relates to tax law the question is in principle similar, but often different in degree. The law that Inyang breached was relatively simple, as was his mistake. A breach of tax law, on the other hand, may well relate to a very complex law that the taxpayer or the taxpayer's advisers have studied at great length. In Inyang's case the crux of the problem could be summed up in a fairly simple question: "Did Inyang believe that he was permitted to use 'physician' so long as he did so in a way that he honestly believed did not mislead?" Inyang's innocence or guilt depended on a simple yes or no answer.

In a tax case, after long study of the facts and law, the state of mind of the taxpayer can vary from certainty that in law the arrangement that is proposed has the tax effect that the taxpayer wishes, to strong suspicion that the arrangement will be overturned if found out, to knowledge that the arrangement is a sham, accompanied by hope that the cloak of the sham will deceive the authorities. The actions of the taxpayer can vary from relentless disclosure of all facts and contracts to the Commissioner to keeping secret transactions that might alert an inspector. Whether the taxpayer is dishonest for the purposes of s 229A should depend on the evidence. In some circumstances, for example, it is submitted that non-disclosure would be evidence tending to cast doubt on a taxpayer's claim of an honest mistake as to the effect of tax law.

2.9 Applying the Test in New Zealand

Take a taxpayer with the third of the states of mind described in the last paragraph. That is, the taxpayer has entered a transaction that he or she knows is a sham, in the hope that the authorities will not discover that it is a sham.[31] The taxpayer is not concerned about evasion because he or she believes that the sham transaction is so complex that the Commissioner will never charge him or her with evasion. The taxpayer believes that the worst that can happen is that the Commissioner will treat the case as one of avoidance. Suppose that the taxpayer believes that it is not dishonest to lodge a tax return that treats the sham transaction as effective, even though the taxpayer has the state of mind that has just been described.

---

30 Ibid, p 803.

31 The meaning of "sham" will be discussed further in Part 2 of this article.

On the basis of *R v Ghosh*,[32] the taxpayer would presumably be convicted of fraud in the UK. In New Zealand the position is uncertain. On one hand, there is the taxpayer's belief that he or she has done nothing dishonest. On the other hand, there is the taxpayer's knowledge of facts that are incompatible with this belief. That is, the taxpayer knows that the tax return will be effective only if it deceives the Commissioner of Inland Revenue; the income is understated, and if the sham transaction is swept away that fact will be clear. One argument is that the taxpayer is using the word "dishonestly" with an incorrect meaning: one cannot claim to be honest at the same time as filing an tax return that one knows to be incorrect. Yet many taxpayers in these circumstances will claim to be honest and we may even believe them when they say that they believe they are honest. In these circumstances, it would be curious if the ratio of *R v Williams*,[33] goes as far as to protect this taxpayer.

It may be that the correct conclusion is that although the New Zealand test of dishonesty is subjective, as held in *R v Williams*, there is at least an element of objectivity when it is a question of the meaning of the word "dishonesty" itself. That is, an accused person probably cannot redefine "dishonesty" according to their subjective opinion of the meaning of the word and then explain that their actions were not dishonest according to that meaning. As Glanville Williams explains:

> "the accused need not have thought of his own conduct in terms of fraud, if he knows all the facts that carry the charge of fraud."[34]

This approach derives some support from *Broom v Police*,[35] where the Court considered the meaning of "corruptly" in the context of s 262 Crimes Act 1961. Citing the *Concise Oxford Dictionary* (6th ed, 1976), Tipping J held that:

> "to act corruptly, a person must (a) have a dishonest purpose and (b) intend to act in a way that can *fairly* be described as morally wicked or depraved."[36]

(Emphasis added.)

The question becomes more difficult as one postulates cases where the taxpayer's state of knowledge is slightly more meritorious. For example, vary the example above so that the taxpayer knows the relevant facts and law that make the impugned transaction a sham, but has not turned his or her mind to them and analysed them. The result is that the taxpayer thinks, to the extent that he or she comes to a conclusion at all, that the case is one of avoidance. Some people would call this state of mind "dishonest". Does it affect the matter that the taxpayer uses a different label for this state of mind? In the current state of the authorities it may be difficult to answer this question, but it is a useful point of reference for a consideration of the examples to be discussed in Part II of this article. To conclude it is helpful to consider a final general point, the effect of non-disclosure by a taxpayer to the

32 [1982] 1 QB 1053. Discussed in section 2.5 of this article.

33 [1985] 1 NZLR 294. See also para 2.6 of this article.

34 Glanville Williams *Criminal Law, the General Part* (2nd ed), London, Stevens & Sons, 1961, p 85, citing *Sayce v Coupe* [1953] 1 QB 6, 7 per Lord Goddard CJ.

35 [1994] 1 NZLR 680.

36 Ibid, p 688.

Commissioner, in the context of the criminal law.

2.10 Non-disclosure

Most of the examples to be discussed in Part II of this article are of cases where a taxpayer undertakes a scheme to minimise tax (to use a neutral term) but does not disclose the scheme to the Commissioner. In these circumstances, the question arises as to whether the fact of non-disclosure can consolidate other evidence against the taxpayer, being evidence that, without more, is not conclusive of a fraudulent intent.

The answer is yes, in some cases the mere fact of non-disclosure can, with other evidence, amount to fraud, even though the other evidence by itself does not necessarily show a fraudulent intent. An example occurred in respect of one of the counts in *R v Gunthorp*.[37] The charge, under s 229A Crimes Act 1961, was against a Mr Hawkins, who was alleged to have used a certain cheque to obtain a pecuniary advantage for himself or for Richardson Camway Ltd, a company that he controlled. The cheque in question was drawn on the bank of Equitycorp Investments Ltd, a member of the Equitycorp group of companies, in favour of another member of the group. Hawkins was chairperson and chief executive of the Equitycorp group. The cheque represented one of a number of transfers of funds that enabled Richardson Camway to refinance a certain debt.

There was a gap in the evidence, in that it could not be conclusively established that Hawkins had "used" the cheque with sufficient knowledge of all the steps in the transaction to fix him with an intent to defraud. However, Hawkins had failed to tell the board of Equitycorp Investments Ltd about the cheque, much less obtain the board's approval. Could this omission amount to fraud?

Counsel for Hawkins argued that failure to disclose could not be fraudulent of itself. For there to be fraud, he said, it had to be shown that Hawkins knew that board approval would not be forthcoming for the transaction. The Court of Appeal examined the transaction and decided that one could not tell from the evidence whether the board would have approved the transaction. The Court went on to say:

> "We do not accept the proposition that there could be fraud only if Mr Hawkins know the Board would not approve. While non-disclosure is not itself enough, concealment of a transaction known to be of a kind that a responsible board would want to examine carefully is . . . The Board may or may not have approved it, it may have approved it with qualifications. Mr Hawkins must have known that. The Judge's finding of dishonesty and fraudulent intent was therefore justified."[38]

In context the Court appears to use "concealment" for stylistic variation; that is, the word does not necessarily imply any steps more active than "non-disclosure". Thus, there is an argument that when the Court's test is transferred to the context of tax cases it can be read: "fraud includes non-disclosure of a scheme known to be of a kind that the Commissioner would want to examine carefully."

This article was submitted for publication on 7 November 1995.

---

[37] 9/6/93, CA46/93, p 60.

[38] Ibid, p 61.

Criminal Law, Tax Evasion, Shams, and Tax Avoidance: Part II — Criminal Law Consequences of Categories of Evasion and Avoidance

JOHN PREBBLE

BA, LLB (Hons) (Auckland); BCL (Oxon); JSD (Cornell); Inner Temple. Barrister, England and Wales and New Zealand. Professor and former Dean of Law, Victoria University of Wellington.

This article discusses the meaning of "evasion" and how it is punished under the Tax Administration Act 1994 and the Crimes Act 1961. It considers the simple case of evasion, where there is an unadorned failure to report assessable income. The article looks at cases where the failure to report income is innocent, then moves to cases where the omission is intentional. Finally looked at are, cases where a failure to report assessable income is associated with a scheme to avoid tax or with a transaction that might have minimised assessable income, or turned revenue into capital, but that did not have those sought-after results.

The first part of this article appeared in Volume 2 New Zealand Journal of Taxation Law and Policy, No 1, March 1996.

3 Evasion in several contexts

3. 1 Introduction

Part I of this article introduced the topic of tax evasion, in particular the topic of fraudulent tax evasion. It endeavoured to explain in a general way how concepts of fraud, as the term is used in criminal law, relate to tax evasion. Part II attempts to relate that general discussion to particular examples, and to show how cases of apparent tax avoidance can sometimes also amount to fraudulent evasion.

3.2 Suppression of income and shams

There are many ways in which tax evasion can occur, and people will adopt all sorts of means to avoid being found out. However, it is helpful to attempt some sort of categorisation of typical cases.

First is the simple failure to disclose income. The shopkeeper who fails to ring up some of her sales on the cash register, and who calculates profits for her tax return on only the sales that she does record, is a familiar example.

Secondly, there is the taxpayer who attempts to cover her tracks by sham transactions. For example, the same shop-keeper may ask a supplier to over-charge her for stock, and to

refund the excess in cash. Orders, invoices, receipts, and bank records will show that stock was bought for, say, $150, but in fact only $100 was ever paid with the intention that the money would stay with the supplier. To the extent that the documents purport to show that another $50 was paid they record a sham transaction that never existed.

Fundamentally, there is no difference between a naked failure to report income and a failure to report income that is accompanied by a sham transaction. Both involve under-reporting of income. The second simply has embroidery added.

On analysis, it may be apparent that a taxpayer has dishonestly under-reported her income, and an accompanying sham may be disregarded. On the other hand, and importantly, a transaction that is a sham may sometimes itself be evidence of dishonesty. Concealing facts from someone who is entitled to inquire into the source or utilisation of money may support a charge of fraud by itself, without evidence of precisely how the accused may have defrauded or attempted to defraud the owner of the money. Thus, Lord Denning said in *Welham v Director of Public Prosecutions:*[1]

> "He [the utterer of forged hire purchase documents] intended to practise a fraud on whomsoever might be called upon to investigate the loans made by the finance companies to the motor dealers. Such a person might be prejudiced in his investigation by the fraud. That is enough to show an intent to defraud."

[1] [1961] AC 103, 134, quoted in *R v Gunthorp* 9/6/93, CA46/93 (NZ), at p 21.

3.3 Transactions ineffective for tax purposes

The relationship between shams and transactions that are ineffective for tax purposes may be illustrated by a hypothetical case that falls somewhere between the facts of *Maney & Sons De Luxe Service Station Ltd v CIR*[2] and *CIR v Dunlop's Wanganui Ltd.*[3]

Suppose that a petrol company agrees to pay a service station a fee of 5 percent of purchases to sell only that company's petrol. Suppose that the petrol company invoices the service station at its regular price, but pays the owner personally in cash by way of refund for the 5 percent reduction. Suppose that the owner treats the 5 percent as a capital receipt, paid for accepting the burden of a trade tie, citing to himself the *Dunlop* case as authority. In computing his assessable income the owner deducts the full invoice price for supplies of petrol.

There are several possible analyses. First, the payment and repayment of the 5 percent may be shams, and the money should be treated as never having left the service station business. In that event, the owner's tax returns are at least arguably evasive, because they suppress income by dint of claiming inflated expenditure.

Alternatively, the payments are not shams, but the scheme does not reduce tax. The reason is that the receipts of cash from the petrol company are not a capital price for a long-term trade tie, but recurring fees for agreeing to trade in only one brand of petrol on a period-to-period basis, as the Court of

[2] [1966] NZLR 41 (CA).

[3] [1970] NZLR 1125 (CA).

Appeal held in the *Maney* case. That is, the taxpayer has wrongly categorised revenue receipts as capital receipts. On this analysis, the taxpayer has under-computed his income by omitting certain revenue receipts.

Whether the taxpayer's returns are evasive may depend on his state of mind. Suppose, to make one possible assumption, that the taxpayer is wholly familiar with the law and knows that there is no doubt that the receipts are revenue items. Suppose that nevertheless he omits them from his income calculations, hoping not to be found out, and planning to mutter "capital in return for a trade tie" if he is discovered. In these circumstances the taxpayer is guilty of evasion, difficult though that may be to prove. As suggested in section 2.8 of this article, it would seem curious if it is a defence for the taxpayer to say, "In these circumstances I do not regard my state of mind as being correctly described as 'dishonest' ".

This hypothetical case illustrates a fundamental point about tax evasion. Evasion may be naked, where there is a simple failure to report income; or it may be cloaked by sham transactions but is, on analysis, still only a simple failure to report income. Alternatively, evasion may be cloaked by transactions that are not shams, but are real and effective, but that nevertheless do not achieve the taxpayer's objective of minimising assessable income.

3.4 Fiscally ineffective transactions — relationship to evasion

This article concentrates principally on the last category of evasion: evasion that is cloaked by transactions that are at least arguably effective in the sense that they create legal rights and obligations, but that are ineffective for the purpose of minimising tax.

The most obvious group within this category are arrangements that are caught by s BB 9 Income Tax Act 1994, which reads in relevant part:

> "Every arrangement made or entered into, . . . shall be absolutely void against the Commissioner for income tax purposes if and to the extent that, directly or indirectly,—
>
> "(a) Its purpose or effect is tax avoidance . . . "

Another group are transactions like the one described in section 3.3 of this article: transactions that as a matter of income tax law do not in fact minimise tax, although the people who participate in them may think that they do. Participants in such transactions may fall into error as to the fiscal effect of the transactions for a number of reasons. Perhaps most often, the reason may be described as a mistake in categorisation: a transaction or a receipt is classified in one way when income tax law would classify it in another.

When people who have undertaken such transactions, or transactions that are caught by s BB 9 of the Act, report their income on the wrongful assumption that the transactions are effective to minimise tax, they are not ipso facto guilty of evasion. But neither should they assume that the presence of a transaction that is caught by s BB 9 but that might not have been, or that might have been categorised as fiscally effective by principles of income tax law, is a kind of immunisation against a finding of evasion. Avoidance and evasion can go hand-in-hand.

Section 4 of this article first considers shams, and then examines a number of categories of transactions that are not shams, but that are nevertheless sometimes ineffective for tax purposes. People who undertake such transactions may be guilty of tax evasion.

Section 5 of the article considers the kinds of circumstances when such evasion may exist. Section 5 considers a taxpayer who has entered a transaction that might minimise his assessable income, but that does not do so. The hypothetical taxpayer reports his income as if the transaction is effective for tax purposes, and does not disclose details of the transaction to the Commissioner.

For ease of exposition, section 5 uses for illustrative purposes a taxpayer who enters an arrangement that is void against the Commissioner pursuant to s BB 9 Income Tax Act, but the points made in section 5 could equally well apply to transactions that are fiscally ineffective for other reasons of tax law, or for reasons of general law.

4 Shams, self-cancelling transactions, context, and mislabelling

4.1 Sham transactions: meaning

In *Snook v London West Riding Investments Ltd* Diplock LJ said that "sham":[4]

> ". means acts done or documents executed by the parties to the 'sham' which are intended by them to give to third parties or to the Court the appearance of creating between the parties legal rights and obligations different from the actual legal rights and obligations (if any) which the parties intend to create. One thing I think, however, is clear in legal principle, morality and the authorities . . . that for acts or documents to be a 'sham', with whatever legal consequences follow from this, all the parties thereto must have a common intention that the acts or documents are not to create the legal rights and obligations which they give the appearance of creating. No unexpressed intentions of a 'shammer' affect the rights of a party whom he deceived."

It does not necessarily follow that a transaction that is a sham is a cloak for tax evasion. Shams are often designed to hide other forms of fraud. On the other hand, and importantly, despite having entered into a transaction that is a sham, a taxpayer may correctly report his income. However, one reason for constructing shams is to hide transactions that evade tax. This article is concerned with such transactions.

4.2 Sham transactions: law

It is commonly stated that, in respect of shams, legal transactions can fall into two categories only. They either are genuine, and effective, and not shams, or they are shams, without effect. As Richardson J said in *Marac Life Assurance Ltd v CIR*, "at common law there is no halfway house between sham and characterisation of the transaction according to the true nature of the legal arrangements actually entered into and carried out."[5] In *Mills v Dowdall*, Richardson J had explained the position more fully:[6]

> "The only exceptions to the principle that the legal consequences turn on the terms of the legal arrangements actually entered into and carried out are: (i) where the essential genuineness of the transaction is challenged and sham is established; and

4 [1967] 1 All ER 518, 528-529.

5 (1986) 9 TRNZ 331, 343 (CA).

6 [1983] NZLR 154, 159 (CA).

(ii) where there is a statutory provision, such as s 99 of the Income Tax Act 1976, [now s BB 9 of the Income Tax Act 1994] mandating a broader or different approach which applies in the circumstances of the particular case. A document may be brushed aside if and to the extent that it is a sham in two situations: (a) where the document does not reflect the true agreement between the parties, in which case the cloak is removed and recognition given to their common intentions . . . ; and (b) where the document was bona fide in inception but the parties have departed from their original agreement while leaving the original documentation to stand unaltered."

4.3 Shams and cognate transactions

People sometimes interpret judicial statements like those just quoted to mean that if a transaction is not a sham, and is consequently genuine, it follows that the transaction cannot be evasive or fraudulent, cannot be impugned for avoidance (except by calling in aid a provision like s BB 9 Income Tax Act 1994, as Richardson J explains), and must be given effect according to its terms and syntax.

Taking a narrow view, some people may contend that that interpretation is strictly correct. But the problem with the interpretation is that it does not allow for four other matters that must be taken into account along with the "sham or genuine, no half-way house" rule. These are the three doctrines of mislabelling, context, and related transactions, and the arguments of logic. Transactions may be genuine, and not shams, but may nevertheless fall foul of the three doctrines; as regards logic, transactions may turn out to have legal effects that are different from the impression conveyed by their terminology, and may turn out to be evasive, even though they are genuine in the sense that they have legal effects that accord with their tenor.

These four approaches — mislabelling, context, related transactions, and logic — overlap with one another to some extent. Sometimes a single case can be analysed pursuant to more than one approach. However, for purposes of exposition it is convenient to consider them separately.

4.4 Mislabelling

As Richardson J said in *Marac Life Assurance Ltd v CIR*: "The nomenclature used by the parties is not decisive and what is crucial is the ascertainment of the legal rights and duties which are actually created by the transaction . . .".[7] Thus, names or labels that parties apply to their transactions are not decisive. In fact, in *Marac* itself the Court of Appeal held that a document labelled "bond" was in fact an insurance contract.

An early, rather simple, case of mislabelling occurred in *Duval & Co Ltd v Federal Commissioner of Taxation*.[8] In addition to rent, a landlord purported to charge tenants "contributions to alterations" to help to defray the cost of improvements to certain rented premises. The landlord treated the "contributions" as receipts of capital. The Court held that the contributions were in fact rent, and were therefore assessable as income.

The holding in the *Duval* case did not entail finding that the leases were shams: simply that the parties had employed an incorrect label for some of the payments that they had contracted for. Nor is it necessary in such a

---

7 (1986) 9 TRNZ 331, 343 (CA).

8 (1933) 2 ATD 293.

case to hold that there is an arrangement that is void against the Commissioner for tax purposes. The Court simply gives the agreement its true legal effect, albeit that this effect is not what the parties had hoped.

A somewhat more complex example is *Leary v Federal Commissioner of Taxation.*[9] This case involved a payment of $10,000 to the Order of St John, most of which was returned to the donor by a series of circular transactions organised by Metropolitan Taxation Services, a scheme promoter. The ultimate disposition of the money was: Order of St John $120; Mr Leary $8,483; fees and profit to MTS or its associates $1,397. The question was whether the $10,000 was a gift, in which case it was deductible by Mr Leary. On the facts, the Court held that the payment was not a gift, because it did not have the characteristics of a gift: "a transfer of a beneficial interest in property by way of benefaction, and an absence of pecuniary or proprietary benefit passing to the transferor by way of return".[10] The label "gift" employed by Mr Leary and the Order of St John was not effective, and Mr Leary's deduction was denied.

The consequences of cases like *Duval* and *Leary* will vary according to the circumstances. Where there has been full disclosure it is simply a matter of correcting the taxpayer's assessment. But concealment, and use of the mislabelling to mislead the Commissioner, could turn such a case into evasion.

4.5 Context

In *Mills v Dowdall,* Richardson J said:[11]

"In characterising the transaction regard is had to surrounding circumstances: not to deny or contradict the written agreement but in order to understand the setting in which it was made and to construe it against that factual background having regard to the genesis and objectively the aim of the transaction."

As with mislabelling, the question of context is highly relevant. One cannot look at a particular element of a transaction, or even of a series of transactions, without taking account of the surrounding circumstances.

4.6 Related transactions: self-cancellation

Again in *Mills v Dowdall,*[12] Richardson J said that there must be "consideration of the whole contractual arrangement and if the transaction is embodied in a series of inter-related agreements they must be considered together and one may be read to explain the others." The inter-related transactions rule is perhaps a sub-rule of the context doctrine that has just been mentioned.

The effect of considering a series of inter-related transactions together was explained by Cooke J (as he then was) in *Mills v Dowdall* as regards the characterisation of a gift. Cooke J's words are equally applicable in other cases if one substitutes "payment" for "gift" and makes consequential changes to his Honour's words:[13]

9 (1980) 32 ALR 221 Full Fed Ct.

10 Ibid, per Brennan J at p 237 line 14.

11 [1983] NZLR 154, 159 line 47 (CA).

12 Ibid, p 159 line 44.

13 Ibid, p 157 line 4.

> "We are concerned with the meaning of 'gift' . . . and a commonsense interpretation is appropriate. If, by prior arrangement with the transferee or otherwise, a transfer of property for a stated consideration were accompanied by an immediate forgiveness by deed of the entire debt, I do not think that the Court would have to shut its eyes to reality. It would not be precluded from holding that the two elements in the transaction were so linked that they should be treated as inseparable; that the transferee was intended never to incur any real liability; and that in truth the property was given to him."

As Cooke J put it,[14] such transactions are self-cancelling. They may not be shams, but in order to determine the effect of a single transaction the Court must determine the overall legal effect of a series or combination of transactions of which it forms part.

Whether transactions must be treated as self-cancelling or as independent is sometimes a difficult decision. A good example is found in *Fitzwilliam v Inland Revenue Commissioners*,[15] a case about a plan to avoid UK capital transfer tax. Steps of the plan included Countess Fitzwilliam giving £2 m to her daughter on 9 January 1980 (step 2) and the daughter giving £2 m to Countess Fitzwilliam on 31 January 1980 (step 4). Lord Templeman held that, "Steps 2 and 4 must be disregarded [for purposes of the plan] . . . The payments are disregarded because they are self-cancelling."[16] On the other hand, Lord Keith and the rest of their Lordships disagreed, and gave independent effect to the two gifts.

In the present context, the importance of the *Fitzwilliam* case lies not in whether on the facts steps 2 and 4 should have been held to be self-cancelling, but as illustrating that the doctrine of disregarding self-cancelling transactions is part of the common law.

4.7 Related and self-cancelling transactions and New Zealand law

Cooke J's words in *Mills v Dowdall* have stood since 1983. But some people question whether the doctrine of self-cancelling transactions is part of New Zealand law. Why is this? One possible reason is that the best-known application of the related self-cancelling transactions rule is in the principle sometimes known as the "doctrine of fiscal nullity" (of which Lord Templeman's approach to the *Fitzwilliam* case is an example) which was expounded by Lord Wilberforce in *W T Ramsay Ltd v Inland Revenue Commissioners*:[17]

> "It is the task of the Court to ascertain the legal nature of any transaction to which it is sought to attach a tax or a tax consequence and if that emerges from a series or combination of transactions, intended to operate as such, it is that series or combination which may be regarded."

Cooke J quoted from the *Ramsay* case in *Mills v Dowdall*. As Cooke J and Richardson J explained in *Mills v Dowdall*, what is called here the related transactions rule is a general rule of law. However, in respect of tax avoidance transactions the related transactions

14 Ibid, p 157 line 20.

15 [1993] 3 All ER 184 (HL).

16 Ibid pp 213-214.

17 [1982] AC 300, 323-324.

rule is codified and expanded in New Zealand by s BB 9 Income Tax Act, and in Australia by the Australian counterpart of s BB 9, namely, Part IVA of the Federal Income Tax Assessment Act. Broadly speaking, these provisions say that arrangements that avoid tax are void as against the Commissioner of Inland Revenue.

The ordinary rule is that where Parliament codifies an area of Judge-made law the legislation supplants the Judge-made law. For that reason, Australian Courts held in *Oakey Abattoir Pty Ltd v Federal Commissioner of Taxation* [18] and *John v Federal Commissioner of Taxation*[19] that the fiscal nullity rule does not apply in respect of tax avoidance transactions: Australia has its statutory rule instead.

Many people believe that the *Oakey Abattoir* and *John* cases apply in New Zealand, and that in New Zealand s BB 9 supplants the fiscal nullity rule. That belief may be correct, though arguably it gives too little weight to Cooke J's words in *Mills v Dowdall.* It is clear from p 157 of the report in that case that Cooke J believed that the fiscal nullity rule applies by common law to tax cases, whether or not there is a provision like s BB 9, and that Cooke J was simply explaining that the related transactions doctrine (of which the fiscal nullity rule is a sub-set) applies generally, in other areas as well as in tax cases.

Be that as it may, suppose that s BB 9 does supplant the fiscal nullity rule in respect of tax *avoidance* cases. The Judge-made related transactions rule remains part the rest of New Zealand law; part of New Zealand tax law in general, and of New Zealand law as to tax *evasion* in particular. Thus in appropriate cases prosecuting authorities have available to them the related transactions doctrine to show (for example and in particular) that self-cancelling transactions may be ignored when a Court is deciding whether a particular pattern of facts amounts to tax evasion.

4.8 *Other judicial authority*

The argument that the related transactions rule applies to tax law in general finds support in *Chadwick v Commissioner of Stamps (NSW),*[20] where the Court held that a payment to X can amount to a disposition to Y. Another example (this time of self-cancelling transactions) is *Re Duthie.*[21]

In *Re Duthie,* the question arose as to whether a father had transferred £1,000 to his son by handing over a cheque for that amount, when the son at the same time gave the father a similar cheque. The son's cheque was purportedly a loan, in return for which the father executed a mortgage in favour of the son. Were the transactions separate (as the father's executor argued), or was the son's loan to the father in fact taken out of the father's gift to the son (as contended by the Commissioner of Inland Revenue)? McGregor J said:[22]

> "While the advance was made by the son it seems an irresistible inference that such advance was made by the son by the utilisation of the sum of £1,000 provided by the [father]. It can be accepted that, for

[18] (1984) 84 ATC 4718 (Fed Ct).

[19] (1989) 166 CLR 417 (H Ct).

[20] (1919) 19 SR (NSW) 39.

[21] [1964] NZLR 1006.

[22] Ibid p 1009.

the fraction of time representing the period from the crediting of the sum of £1,000 to the son's account to the debiting of the second cheque of £1,000 to this account, the sum of £1,000 was under the son's control or within his enjoyment. Nevertheless, it seems to me that it is an irresistible inference that the arrangement between the parties was that the [father] would give to his son a mortgage for £1,000, the moneys thereby secured to be provided by the gift of £1,000."

The argument that is advanced here is not that if transactions are held to be self-cancelling they must together amount to tax evasion. Rather, the argument is that when a Court is deciding whether there has been evasion, in a suitable case one of the steps in the Court's reasoning may be to find that self-cancelling transactions are involved. Further steps in the Court's reasoning will examine the issues of whether relevant parties knew or should have known that the transactions were self-cancelling and whether this knowledge, in the circumstances of the case, should lead the Court to determine that there was dishonesty.

4.9 Logic

Analysis reveals that the statement "the transactions are effective, not shams, therefore there is no tax evasion or fraud", involves a non-sequitur. Fraud is often the bedfellow of contracts that are effective in one or more respects. For example, a contract of sale between a victim and a rogue that is voidable at the instance of the victim for mistake because of the fraud of the rogue (but not void) is effective at law to pass sufficient title to the rogue for him or her to give a good title to a third party purchaser. That situation has been continued, and probably extended, by s 8 Contractual Mistakes Act 1977.

Take a simple example from the tax area. Suppose that the landlord/taxpayer in *Duval v Federal Commissioner of Taxation*[23] well knows that, by law, tenants' "contributions to improvements" are assessable as rent. Suppose that the taxpayer nevertheless drafts his leases to require rent, and adds collateral contracts for his tenants to pay "contributions to improvements". In his income tax returns the taxpayer omits any mention of the "contributions", and when asked by a tax inspector, "Are you still collecting contributions to improvements?" the taxpayer answers "No".

In these circumstances there can be no doubt that the suppression of the "contributions" is evasion, and not avoidance. And yet payments under the collateral contracts will be effective to transfer ownership of the "contributions" from the tenants to the landlord/taxpayer.

4.10 Illegality of contracts

The effectiveness of a contract to transfer property from one person to another where the contract is part of an arrangement that is contrived to evade tax requires some discussion. Where one party tries to enforce the contract against the other he or she will be unsuccessful at common law, though in New Zealand the Illegal Contracts Act 1970 affords relief pursuant to the discretion of the Court. However, an illegal contract is effective to pass title to property when the contract has been performed.

The distinction between trying to enforce an illegal contract and taking a benefit under it by virtue of its performance is illustrated by *Napier v National Business Agency Ltd.*[24]

---

[23] (1933) 2 ATD 293. See section 4.4 supra.

[24] [1951] 2 All ER 264 (CA).

Napier was employed by NBA at £13 a week plus £6 a week reimbursement for expenses. Both parties knew that Napier's expenses would be negligible. Categorising the £6 as expenses was simply a tactic to evade tax on that sum. Napier was sacked, but because the contract was illegal he was unable to get damages for lack of notice.

There was no question in the case of NBA trying to recover from Napier the salary and "expenses" that it had paid to him in the months up to his dismissal. Any such endeavour would have failed, because the contract, although illegal, was by common law effective to transfer property in the money that NBA paid to Napier.

5 Evasion in the context of schemes to avoid tax

5. 1 *Continuum*

What actions by taxpayers that abstract money from the Inland Revenue Department, or that fail to pay money to the Department, amount to evasion? In responding to this question, one can consider courses of action by taxpayers as if on a continuum. At one end there are actions that are clearly evasion — deliberate under-reporting of turnover by a retailer is an example. At the other end there is an avoidance transaction that is (let it be supposed) void by s BB 9, but that is fully disclosed to the Commissioner. Both of these examples may cause loss to the Department, but the second is not evasion because it lacks the element of concealment that was identified by Lord Templeman (p 3 of part 1), and is for that reason not dishonest.

Between the ends of the continuum there are many examples of many different types of activity. For the moment, focus on avoidance schemes that the taxpayer does not disclose to the Commissioner. Now, classify those schemes according to whether they are, or are not, struck down by s BB 9, or by some other anti-avoidance rule, or by a general principle of tax law.

5.2 *Schemes that are effective for tax purposes*

If the schemes are successful, and survive challenge under the Act or under general principles of law, logically there can be no evasion. The reason is that there can be no loss to the Department. Similarly, it would seem that there can be no offence under s 229A(b) Crimes Act 1961 (See J Preb PRTI p3, 7ff). The reason is not necessarily that the taxpayer's motive was not dishonest, nor that he did not intend to abstract money from the revenue. The reason is that the document that the taxpayer used was not capable of producing a benefit. Accordingly, even if the taxpayer's motive was dishonest he cannot be convicted of evasion. It may be that a taxpayer in dishonestly using any such document could be guilty of an attempt, but this article will not pursue that question.

Just because a scheme is effective for tax purposes it does not necessarily follow that the taxpayer cannot be guilty of any form of fraud. The taxpayer or his or her advisers are conceivably guilty of conspiracy to defraud if their actions amount to doing a lawful action by unlawful means.

An example is a scheme devised for a company that involves (a) minimising tax liability, and (b) ensuring that, if the scheme fails, there will be no money left in the company for the Commissioner to collect. Even if step (a) is eventually determined to be tax effective, step (b) can amount to conspiracy to defraud, as seen in the case of *O'Donovan v Vereker*.[25]

---

[25] (1987) 76 ALR 97 Full Fed Ct.

This paper leaves to one side discussion of conspiracy such as was alleged to have occurred in *O'Donovan v Vereker*, to concentrate on the substantive offence of criminal evasion itself.

5.3 Schemes that are ineffective for tax purposes

A tax minimisation scheme may be ineffective for tax purposes for many reasons. The most obvious is that it is void against the Commissioner under s BB 9. Other reasons include breach of specific anti-avoidance rules and misunderstanding or misapplication of general principles of law. Some relevant general principles and concepts are classified in section 4.3 of this article: shams, mislabelling, interpretation by reference to context, self-cancelling transactions, and simple logic in legal reasoning.

Postulate, for present purposes, schemes that are certainly ineffective, in the sense that if they went before the Courts they would be struck down for one or more of the reasons mentioned. But bear in mind that the facts of any particular case (that is, the case of a particular taxpayer in respect of a particular year) have not yet gone before the Courts. In advance, therefore, one cannot say with scientific certainty that the scheme involved in that case would not survive. Postulate further, that the taxpayer lodges his return on the basis that the scheme is effective, but that he does not disclose the details of the scheme to the Commissioner.

On these facts, all the ingredients of an offence under s 229A(b) of the Crimes Act exist, except possibly for dishonesty. In this connection, is it a defence for the taxpayer to say, "Granted, I did not disclose details of the scheme to the Commissioner. However, I did not know with scientific certainty what I now do know, that the scheme is ineffective for tax purposes. Therefore I was not dishonest and I am not guilty". This might be described as the "blind eye" defence.

5.4 The blind eye defence

Clearly enough, the blind eye defence will not be successful in all circumstances. One example is a tax scheme that is so hopeless that it is obviously ineffective. For example, suppose a solicitor says to a client, "Don't pay me, pay my wife", and the client complies. If the solicitor lodges his tax return on the basis that the fee is not assessable he cannot be in any better position than his colleague who simply, and intentionally, omits a fee from his tax computations. Both are guilty of evasion.

Another example is the taxpayer who repeats a plan that has failed in the Courts. Suppose a taxpayer were to re-run the paddock trust scheme that the Privy Council struck down in *Mangin v CIR*.[26] Suppose that the taxpayer is aware of the *Mangin* case, but he reports his income on the basis that the scheme is effective to reduce his taxable income. He does not disclose the scheme to the Commissioner. When charged with evasion, can the taxpayer successfully defend himself by saying, "Granted, the Privy Council struck down a paddock trust in *Mangin*, but I intend to fight my case all the way to the Privy Council again, and hope to have the decision reversed"? The answer must be no; the taxpayer is guilty of evasion.

5.5 More meritorious cases

Take a case at the other end of the scale. Suppose the taxpayer in *Challenge Corp Ltd v*

26 [1971] NZLR 591

CIR[27] had lodged its return without disclosing the scheme that it had adopted. Assume that the failure to disclose was deliberate. (Of course, Challenge did disclose the scheme; the company even asked for a ruling before proceeding.) Eventually, it was discovered that the scheme was ineffective, with the result that it became apparent that Challenge's income had been understated in its return. If a company in the position of Challenge had lodged its return, deliberately without disclosing the scheme, would it have been guilty of an offence under s 229A(b) Crimes Act? Assume that the reason for concealment is that, while the company thinks that its scheme is effective, it would prefer not to have the matter tested, in case it is proved wrong.

There are two possible responses. First, the notional company would not be guilty, because there was so much uncertainty as to the effectiveness of the scheme that it was not dishonest for the company to proceed, even though the company did not disclose relevant facts to the Commissioner. After all, the Court of Appeal thought that the *Challenge* scheme was effective, so one cannot complain that the notional taxpayer acted on the same theory. Alternatively, the company is guilty, because it concealed relevant information at the same time as using a document to obtain a benefit for itself.

If the first response is correct it means that whether one is guilty of evasion in circumstances of this kind is a matter of degree: the more improbable one's concealed avoidance scheme, the more likely it is that one is guilty of evasion in terms of s 229A(b) Crimes Act. The second response is logically more robust: guilt does not depend on matters of degree, but on concealment that cannot be explained by an honest state of mind.

[27] [1986] 2 NZLR 513, 561; (1986) 10 TRNZ 161 (PC).

5.6 Judicial authority

Both responses derive some support from the case of *O'Donovan v Vereker*,[28] though from different judgments. Pincus J said:[29]

> "I cannot, with respect, accept [Northrop J's] view that the degree of probability of success of the NIPAG [tax avoidance] scheme is irrelevant in considering fraudulent intention. In my opinion the magistrate [in deciding whether to send the case for trial] may properly consider the question whether there was a 'clear risk that the CYP companies [companies with current year profits seeking a tax shelter] would be assessed', as Fox J says in his reasons [in the third judgment in the case], and, to put the matter more generally, that he may properly consider the degree of probability, as it appeared to the participants in the scheme, that the companies would be correctly assessed."

On the other hand, Northrop J said that even "[a]n honest belief that the scheme was effective . . . does not prevent the finding of a criminal intent."[30]

The *Vereker* case involved facts slightly different from those postulated in the example now under discussion. The prosecutor alleged that the avoidance scheme attempted both (a) to minimise tax, and (b) to contrive that the taxpayer would have no money for tax if objective (a) failed. Thus, Northrop J is saying that it does not matter whether the taxpayer honestly believes objective (a) will succeed, for objective (b) is fraudulent on its own.

[28] (1987) 76 ALR 97 Full Fed Ct 7.

[29] Ibid p 128.

[30] Ibid p 116.

On the other hand, Pincus J is saying that in assessing whether the taxpayer's actions in pursuit of objective (b) are fraudulent it is relevant to consider his state of mind in respect of objective (a). Did the taxpayer honestly believe that there was a high degree of probability that the minimisation scheme would work? (In which case he is probably not guilty of fraud.) Or must the taxpayer, with the level of knowledge that he had and that must be imputed to him, have entertained serious misgivings about the robustness of the minimisation scheme if the Commissioner should hear of it? (In which case he probably is guilty.)

5.7 The Vereker case and prosecutions under s 229A(b) Crimes Act

There is a close fit between the facts in the *Vereker* case and the kind of prosecution under s 229A(b) Crimes Act that is now under discussion. In the *Vereker* case, the question was: did the protagonists adopt measures to ensure that there would be no assets from which to pay tax whether or not the minimisation scheme was effective? In a s 229A(b) case that involves a concealed avoidance scheme, the corresponding question is: did the taxpayer use a document (perhaps an invoice) to get a benefit, concealing facts that might or might not have disentitled him to the benefit? The asset stripping element in the *Vereker* case corresponds to the concealment element in the s 229A(b) case. Both are calculated to prevent the Commissioner from collecting tax (though in different ways) if it should be that the minimisation scheme does not work and there is assessable income to be taxed.

It follows that the Judges' opinions in the *Vereker* case can be directly applicable to s 229A(b) prosecutions. If Pincus J is correct, the probability of success of the minimisation scheme, as seen by the taxpayer, is relevant to guilt. If Northrop J is right, the taxpayer can be guilty even if he was confident in the probability of success of the scheme.

Whichever way one looks at it, and Pincus J's approach favours the taxpayer who is prosecuted, *O'Donovan v Vereker* is authority for the proposition that if taxpayers lodge returns on the basis of the assumed effectiveness of non-disclosed avoidance schemes, then some of those taxpayers will probably be guilty of offences under s 229A(b) Crimes Act.

5.8 Significance of R v Forsyth

One sequel to *O'Donovan v Vereker* was *R v Forsyth*.[31] Forsyth was a silk who had been engaged to advise other defendants in the *Vereker* case at the planning stage. He was prosecuted for various offences that related to fraud, including conspiracy. The burden of the allegations against Forsyth was that he had told potential customers of promoters of an avoidance scheme that the scheme was lawful when he knew that it was not. Procedurally, the case started with a preliminary hearing before a Magistrate. The Magistrate held that there was a case for Forsyth to answer, and sent him for trial.

Forsyth applied to Jackson J in the Federal Court for review of the Magistrate' s decision. Jackson J's decision is reported as *Vereker v Rodda, Forsyth v Rodda*.[32] *O'Donovan v Vereker* concerned appeals to the full Federal Court against Jackson J's decision in *Vereker v Rodda, Forsyth v Rodda*. On appeal, Forsyth was not successful, and his trial later proceeded before Hampel J and a jury.

---

[31] (1990) 20 ATR 1818.

[32] (1987) 72 ALR 47.

At the trial, on the completion of the case for the prosecution, counsel for Forsyth argued that a properly instructed jury could not convict Forsyth on the evidence presented by the prosecution, and that the jury should be directed to acquit him. This argument was successful, and Hampel J so directed. The report in *R v Forsyth* contains Hampel J's reasons. In short, the reasons were that an essential connection between Forsyth and the facts of the case had not been established.

In the present discussion, the importance of *R v Forsyth* is not that a prosecution for fraud that is perpetrated in the context of an avoidance transaction cannot be prosecuted or cannot be fraud. On the contrary, it is basic to the decision of the Full Federal Court in *O'Donovan v Vereker* that actions of individuals in the context of an avoidance transaction *can* amount to fraud. However, on the facts as presented at the trial a case was not made out against Forsyth. Hampel J's judgment in *R v Forsyth* does not disturb the statements of principle by the full Federal Court in *O'Donovan v Vereker*. Indeed, they could not be disturbed, because Forsyth's trial took place in an inferior Court.

5.9 Relevance of disclosure requirements of the Tax Administration Act

The provision in the Tax Administration Act 1994 that requires taxpayers to file returns is s 33. The section does not in explicit terms require taxpayers who have undertaken avoidance schemes that may affect their assessable income to disclose details of the schemes. It requires simply that a taxpayer should "for the purposes of the assessment and levy of income tax furnish to the Commissioner in each year a return or returns in the prescribed form or forms setting forth a complete statement of all the assessable income derived by the taxpayer during the preceding year . . .". Is it a defence to a charge under s 229A(b) Crimes Act that the taxpayer was not required by s 33 Tax Administration Act to disclose details of avoidance arrangements? The answer is no, it is not a defence, for two independent reasons.

5.10 Compliance with s 33 Tax Administration Act not directly relevant

First, a prosecution under s 229A is under the Crimes Act. The fact that the defendant has complied with obligations under another Act (if he has so complied) is logically not directly relevant to the offence that is the subject of the prosecution. Section 229A describes what is in a sense an amorphous, substantive offence. To paraphrase, the ingredients of the offence are:

(a) using a document capable of being used to obtain a benefit,

(b) dishonestly,

(c) for the purpose of obtaining a benefit.

Logically, formal, or even substantive, compliance with some other rule of law cannot determine whether these ingredients are present.

There may be an exception to the statement in the last paragraph. Compliance with the formal requirements of tax law could be relevant to the issue of dishonesty. Suppose a person who is accused of an offence against s 229A Crimes Act advances the defence that he honestly believed (1) that what he had done complied with s 33 Tax Administration Act, and (2) he honestly believed that compliance with s 33 was all that the law required of him. Subject to the discussion of mistake of law in

section 2.7 (See Part 1 of this article)[33], there is some possibility that this plea, if believed, is an effective defence.

But is there a flaw in this argument? The second "honest belief" is surely only a short way of saying that the defendant believed that his actions could not be dishonest if he complied with s 33 Tax Administration Act. That belief appears to be a belief that relates to the meaning of "with intent to defraud" in s 229A Crimes Act. Such a belief, even if an honest mistake, cannot serve as a defence to a charge under s 229A.[34]

5.11 Non compliance with s 33 in any event

Secondly, in the kind of case that is now being considered there would not in any event be compliance with s 33 Tax Administration Act. The reasons are as follows. The hypothesis under discussion entails a case where a taxpayer has put into effect an avoidance arrangement. The arrangement is in fact ineffective, and void against the Commissioner. The taxpayer may not know this fact with mathematical certainty when she lodges her return of income, but, according to the hypothesis, the ineffectiveness of the arrangement for tax purposes is nevertheless a fact.

Strictly speaking, it is not relevant to the present discussion *how* we know that the arrangement is ineffective for tax purposes. One possibility is that ineffectiveness becomes apparent ex post facto, perhaps as a result of litigation. Bear in mind that if an arrangement is void for tax purposes, s BB 9, or general rules of law, operate immediately, from the time of putting the arrangement into effect.[35] In other words, if an arrangement is void, its ineffectiveness is a fact from the time of its execution; fiscal ineffectiveness does not wait to become a fact until the Commissioner alleges it or a Court determines it. This position is, incidentally, different from the rule under the current Part IVA of the Australian Income Tax Assessment Act 1936, which does not bite until the Commissioner forms an opinion on an impugned arrangement.

To return to s 33 Tax Administration Act, consider a case where a taxpayer might have, for example, income of $100,000. But he has undertaken an avoidance arrangement that purports to reduce that income by $40,000 to $60,000. In his tax return the taxpayer states his income as $60,000 and does not disclose details of the avoidance arrangement. The taxpayer is prosecuted for fraud under s 229A. If the avoidance arrangement is in fact effective it is probably not possible to obtain a conviction, because the taxpayer's income is in fact only $60,000 and what he did could not cause loss to the Commissioner. One says "probably" bearing in mind the arguable possibility of a conviction for attempt or a conviction for trying to deceive the Commissioner, reckless as to whether one's tax return was or was not correct in law. Offences with some similarity to these are discussed in *O'Donovan v Vereker*[36] but are not part of the subject-matter of this article.

If the arrangement is void for tax purposes it has no effect on the taxpayer's income, which

---

[33] See Prebble "Criminal Law, Tax Evasion, Shams, and Tax Avoidance Part 1" *NZ Journal of Taxation Law and Policy*, vol 2, No 1, p11ff.

[34] Crimes Act 1961 s 25, and see the discussion at Part I section 2.9, supra.

[35] *Newton v Federal Commissioner of Taxation* [1958] AC 450 (PC).

[36] (1987) 76 ALR 97 Full Fed Ct.

is $100,000, not $60,000. It follows that the taxpayer's return does not comply with s 33 of the Tax Administration Act because he has not set forth a complete statement of all assessable income: $40,000 is missing. It is no answer for the taxpayer to say, "I was mistaken. I acted on the assumption that the arrangement was effective", or, even, "I thought the arrangement was effective". Section 33 demands substantive and formal accuracy. Failure to comply with s 33 is not excused by a mistake.

5.12 Conclusion

Section 5 of this article has attempted to show how someone may be guilty of tax evasion, and of an offence under s 229A Crimes Act if:

- She undertakes a scheme that is void for tax purposes by virtue of s BB 9 of the Income Tax Act;
- She fails to disclose the scheme to the Commissioner of Inland Revenue;
- She under-declares her income because she returns it on the basis that the scheme is effective, although it is in fact void.

It is not argued that these ingredients by themselves will necessarily be sufficient in all cases. One could reach that conclusion only after (as a start) choosing the approach of Northrop J over that of Pincus J as these approaches are discussed in section 5.6 of this article. What is argued, however, is that the analysis that has been illustrated by reference to schemes that are void pursuant to s BB 9 is applicable equally to transactions that for tax purposes fall foul of the general law. Thus, for example, someone may be guilty of tax evasion where:

- She labels a component of a transaction that she undertakes with a name that would have the effect of minimising tax (for example, "gift" rather than "payment");
- The label is wrong;
- She fails to disclose the transaction to the Commissioner of Inland Revenue in detail sufficient for the Commissioner to decide whether the label is correct;
- She under-declares her income because she returns it on the basis that the label is right, although it is in fact wrong.

This article was accepted for publication on 7 November 1995

The Dependence of Morality On Law†

TONY HONORÉ*

I *Morality and Law*

The relation of morality to law forms a central part of legal philosophy. It is a complex topic; but over the last forty-five years much has been done to unravel its complexities. Though many have contributed to this process, for clarity and candour Hart's writings still stand out, as they do for the depth and range of the debate they have provoked. This contribution to the series in his honour will not however directly assess his views or those of his critics. Instead it will proceed from a different, though in some ways traditional, point of view. It will be in parts more political than philosophical.

Much writing about morality and law rests on an unstated premise. The premise is that we can determine what morality requires independently of the law of a community. What is right or wrong, good or bad, can be settled by rational argument without recourse to the formal institutions that make up a society's laws and legal system. If morality is universal, the same for everyone, that is still more clearly the case. Laws vary from one society to another, morality does not. Arguments of principle can settle what conduct is morally right or wrong, good or bad. This autonomous morality then provides a blueprint against which to judge laws and legal systems. Do they come up to scratch when set against the blueprint?

A whole series of questions then emerge. Are moral requirements, or some of them, necessarily part of a society's law? How far should they be made part of it? If a law violates these requirements, is it invalid? Are legal obligations 'obligations' in the sense in which moral obligations clearly are? Or are they obligations in a different sense, and with a different force? Over the forty-five years since Hart reopened the debate about law and morality at Harvard in 1957,[1] these and cognate issues have been keenly contested. On all of them the last word remains to be said.

But is the underlying premise correct? Can what morality requires be settled independently of law? Or is the picture of morality and law as model and copy in some ways misleading? Without reverting to the view which saw morality as a

† A version of the eighth H. L. A. Hart lecture sponsored by the Tanner Trust and delivered in Oxford on 12 May 1992. I am grateful to Joseph Raz and John Gardner for commenting on an earlier version and to Ronald Dworkin who presided over and others who attended the seminar on 13 May which followed the lecture.

* All Souls College, Oxford.

[1] 'Positivism and the Separation of Law and Morals': (1958) 71 *Harv LR* 593–672.

© Oxford University Press 1993 Oxford Journal of Legal Studies Vol 13, No 1

form of legislation (something like the Ten Commandments), I shall put the case for thinking that in any complex society a viable morality must have a legal component. It is incomplete without law and cannot be taken seriously unless filled out by law. Some obligations can be spelled out only by law and in certain moral conflicts law plays a special role. So some laws at least relate to morality in two ways: they form part of it and at the same time are open to moral criticism.

How are we to understand morality, a term with rather uncertain limits? It is concerned with conduct that has a significant impact on other people, and perhaps also animals, individually or collectively, and with the restraints on behaviour that we should accept because of this. Moral criticism assesses behaviour in the light of its impact on others. It excludes purely self-regarding behaviour. Moreover, since we live in groups and communities, and belong to states and other political entities, the central core of morality is concerned with how to co-exist and co-operate with others. The core of morality is, in a broad sense, political.

Moral reasons for conduct can be contrasted with technical or instrumental reasons. Not every reason for behaving in one way rather than another is moral. There can be reasons for acting that belong to a specialized activity, for example reasons internal to chess or commerce. These focus attention on a particular aim, like winning the game or making a profit. If the aim has value, and is appropriate in the circumstances (which depends on moral considerations), what we do is assessed from the point of view of its tendency to advance the aim. No further moral question arises. Much legal reasoning seems in this sense to be instrumental. It is concerned, for example, with keeping the law consistent with authority, it being assumed that there is value in doing so. But the decision to adopt the technical chess-playing, profit-making or systematic legal point of view can itself be a subject of moral criticism. It is not always appropriate to play chess or maximize profit or follow authority. In reasoning about law, as about anything else, a great many decisions are not technically determined; and sometimes, even when a technical point of view could be adopted, it should for moral reasons be rejected.

On this view of morality, can there be a morality without a legal component? Certainly in communities which do not have formal institutions, there will be a co-operative morality independent of law. In these (perhaps imaginary) simpler societies there are no laws, and morality is made determinate mainly by convention. But in complex societies it is doubtful whether morality could dispense with law. So an alternative view is worth considering. According to this the resources of morality are commonly overestimated. Morality on its own is incomplete and cannot provide a viable guide to what we are required to do in particular situations. Moreover there are some moral conflicts that morality cannot on its own resolve. For moral thinking to guide the conduct of people living in communities and belonging to political entities, as we all do, it needs to be filled out by various means, including formal institutions. The main formal institutions are laws and legal systems.

To say this is not just to assert that law influences moral opinion, as moral opinion influences law, though their mutual influence is undeniable. Moral argument against sex or race discrimination leads to laws banning discrimination, and these laws in turn reinforce, or sometimes erode, the opinion that discrimination is wrong. Nor is the point being made that, given the shortcomings of human nature, laws are needed to give an incentive to conform to moral requirements. Laws reassure those who are disposed to conform that it is worth doing so and deter those who are disposed not to conform. That too is undisputed, and represents one way in which law is indispensable to a viable morality. But it shows only that some parts of morality cannot be enforced without law, not that what is right or wrong depends on what the law is. Again, the argument is not that morality embodies law because prima facie there is a duty to obey the law. Some of those (including myself) who think that there is such a duty base it on the idea that members of a community are morally committed to its institutions. No such general commitment is here presupposed, only that there is a commitment to a morality of which the core is co-operative. Nor am I directly concerned with the importance of legal casuistry for moral thought, a topic developed in an illuminating way by writers such as Atiyah and Wertheimer in their discussions of promising[2] and coercion.[3]

All these connections are worth exploring. They point to ways in which laws influence moral opinion or suggest how morality might be developed or its viability guaranteed. This essay is however directed to two other topics: law's role in regard to gaps in morality and to moral conflicts. My argument is that over a range of cases there can be no way of determining the right course of action without a legal component. Even a society of well-disposed angels, uniformly anxious to do right, needs a system of laws in order to know the right thing to do.

On this view law is part of the morality of any complex society. It follows that moral criticism directed against those laws that are part of its morality must take account of an awkward fact. If those laws are not morally compelling the society will to that extent lack a viable morality. The picture of morality as a blueprint and law as a structure put up according to or in disregard of it is then seen to be misleading. Morality is more like an outline from which details are missing. Laws, along with conventions, fill many of these in. As the legislator, judge, or official does this the details are incorporated in the overall plan, unless they diverge so far as to be incompatible with it. But we should be slow to reach the conclusion that the details do not fit the blueprint. If we do the building may have to be pulled down.

II *Morality as an Incomplete Guide*

I begin with morality as an incomplete guide to conduct. Morality, in the sense already explained, that of a co-operative morality which guides and assesses

[2] P. S. Atiyah, *The Rise and Fall of Freedom of Contract* (1979); *Promises, Morals and Authority* (1981).
[3] A. Wertheimer, *Coercion* (1987) 10–14 and passim.

conduct from the point of view of its impact on others, can refer to a number of related things. There are the mores of a community, practices which consist in or are deducible from observed behaviour. These may not be uniform. Then there are the moral opinions of a community, what people in it regard as right or wrong, good or bad. These also may diverge and may not coincide with its mores. Finally, there are moral views which stand up to rational scrutiny. These in turn may not agree either with the dominant mores or moral opinions of any actual community. Take abortion. One can distinguish the practices—how many pregnant women have or refuse abortions—from whether those practices are regarded as right or wrong by the community in which they occur, or part of it. Both are distinct from the further question whether abortion really is right or wrong, or sometimes right and sometimes wrong; or, for those who think there is no such thing as objective right or wrong, whether it is rationally defensible or indefensible.

Though morality in all three forms is to some extent incomplete, it is the third type of morality, critical morality, that most clearly calls for a legal supplement. I assume that critical morality is meant to be taken seriously. Thinking about right and wrong is an activity whose aim is not just to find out what would be right in an ideal society. It must include the aim of guiding behaviour in real life. Given this aim critical morality is incomplete as a guide to conduct as it affects others. The reason is that rational moral criticism has to be based on values that can be defended as worth pursuing. But these values are so general that they do not by themselves determine how people should behave in a given instance. We can seldom proceed by a process of deduction from the values to the required behaviour.

Early in the history of philosophy Aristotle made this simple point. In a well-known passage near the beginning of the Nicomachean Ethics[4] he says that politics, a term which in his use includes morality, admits only a limited degree of precision. The subjects it studies, such as noble conduct and justice, are variable and uncertain. They seem to exist only by convention or law (*nomos*) and not in nature. Whatever his own view[5] (and as I am not sure what it was I do not claim his support) at the outset of his course of lectures he invites his students to think about the problem which later, with Aquinas,[6] is called that of determination (*determinatio*). This is the problem of making determinate a value such as noble conduct or justice. The process by which this occurs is best called 'determination' rather than concretization, an ugly term, or implementation, a misleading one. Determination entails, for example, that some mechanism be found for settling what justice requires in a given situation. The same is true with the necessary adjustment for all the other values reflected in morality.

The topic of determination became prominent in the context of natural law

[4] 1.iii.1–2; cf 5.i.14.

[5] His view is not altogether clear (see 1.iii.4; 5.vii.1–5). He may have thought that these virtues exist only in a combination of nature and *nomos* (the view put forward in this essay) or that they can exist separately in each but in a different sense.

[6] Summ. Theol. 1–11 q.95 a 2c, on which see J. Finnis, *Natural Law and Natural Rights* (1980) 284–9.

theory. But its importance does not depend on the premiss that the underlying values made determinate are self-evident or part of the nature of things rather than chosen, or partly chosen, by human beings. Every morality rests on some general value or values, some goods regarded as worth pursuing, even if their relation to the fabric of the universe is problematic. Moreover the present discussion is not directed, as are some arguments from natural law, to showing that many legal provisions are morally neutral, like the rule of the road, since the law could have been different and yet remained morally sound. While the choice between different determinations (for example different methods of traffic control) may sometimes be morally neutral, it does not follow that the choice between having determinate rules and not having them is morally indifferent. That there should be a determination is often necessary to the existence of a viable morality. If a busy town has no traffic regulations its moral standing, so far as avoiding injury to others is concerned, is not on a level with that of a busy city with regulations which might have been different. In one drivers and pedestrians know within limits how to behave. In the other people must decide for themselves how best to avoid threatened collisions as they move around. Mistakes are multiplied, and the duty to avoid harming others will not be carried out as well as it would otherwise have been, given the same degree of goodwill.

The need for determinants of morality is particularly clear as regards obligations owed by members of a community to their community. Taxation affords a good example. According to most people's moral outlook members of a community should make a contribution to the expense of meeting collective needs. A morality which denied this would hardly count as co-operative. In a monetary economy the contribution has to be mainly in money, and takes the form of paying taxes. So members of a community have in principle a moral obligation to pay taxes. But this obligation is incomplete or, if one prefers, inchoate, apart from law. It has no real content until the amount or rate of tax is fixed by an institutional decision, by law. What amounts to a reasonable contribution is not otherwise determinable, since what is required is a co-ordinated scheme which can be defended as fair not merely in the aggregate amount it raises but in its distribution. Taxpayers cannot settle it for themselves, as people can within limits settle for themselves, say, the proper way of showing respect for the feelings of others. Apart from law no one has a moral obligation to pay any particular amount of tax. An obligation to pay an indeterminate sum is not an effective obligation; it requires only a disposition, not an action. So, apart from law no one has an effective obligation to pay tax. If there were a society in which morality was taken as a sufficient guide to conduct apart from law it would therefore not be viable. It would grind to a halt and disintegrate for lack of resources. For this crucial moral and political obligation, vital to the life of a complex community, morality depends on law in the sense that to create an effective obligation it must have recourse to law.

It does not follow that the level and incidence of taxation are immune from moral criticism. The law is not conclusive of the obligation to pay, since the

amount or distribution of the tax may be open to criticism. Whether taxes should be uniform or progressive, direct or indirect, as low as possible or as high as necessary depends on considerations of justice as well as on ease of assessment and efficiency in collection. But justice cannot by itself create an obligation to pay any particular amount of tax. So even if the tax is open to criticism, the criticism has to be viewed in the light of the fact that there is no way of fulfilling the obligation to help support the community apart from paying tax. The poll tax currently in force in Britain was, at least in its original form, widely thought to be not merely ill-designed but unjust. But to argue, as some did, that there was therefore no obligation to pay it was to dispense people from their moral obligation of contributing to the cost of running their local community. That is not an argument lightly to be accepted. There may be a choice of evils: injustice on the one hand, default on the other. For those who have to make the choice the legal determination has moral weight, not just because it is institutional, but because it constitutes the only way in which the moral obligation to support one's community can be given effect. For those individuals who feel strongly in a case like this to make a voluntary contribution instead would not be equivalent. Their act would be one of generosity, not the fulfilment of an obligation.

But even if a determination is needed to make some moral obligations effective must it take the form of law? Morality, some think, has the resources to make its own determinations. Or they can be made in other ways, without recourse to law. Four possible mechanisms for this will be discussed: subsumption, calculation, convention and command. The first can be briefly passed over. Sometimes a particular instance can be subsumed under a wider principle. If, for example, morality requires one to keep promises, then a person who makes a promise must do what he promised. The moral principle, combined with the facts, gives rise to an obligation. But in many situations there is no such simple connection with a moral principle, because the principle involved, for example that we should act with due concern for others, is not specific enough.

The second possible determinant, calculation, is an internal resource of morality. One may argue that, while some moral theories do not yield determinate obligations, others enable us to fix our obligations without recourse to law. If morality consists in performing duties and respecting the rights of others, an institutional determination is needed to fix these duties and rights. But if morality consists in maximizing certain consequences, or indeed in maximizing any values at all, people can calculate for themselves what conduct will maximize the appropriate values. In this way, for example, assuming agreement on the consequences or other values to be maximized, everyone can discover what amount of tax he should pay.

But the suggested calculation is impracticable, at least as regards obligations owed by members of a community to that community. To calculate one's proper tax liability, each person would need to know the total sum required and the contribution that everyone else ought on a similar calculation to make. The reason why individual tax liability cannot be calculated in isolation is that a co-

operative morality involves some commitment to distributive justice. But it is not feasible to make these complex calculations, quite apart from the well known difficulties of comparing and quantifying values. Even if it were, the taxpayer could not be relied on to make them without self-preference. Self-assessment to a tax such as income tax is certainly possible when the rate of tax is fixed by law. But if each taxpayer is to fix the rate for himself or herself there will be no agreement on the rate. The sum of the differing assessments will not meet the total need.[7]

Individual calculation cannot in practice take the place of law as a determinant under a maximizing morality any more than under one based on duties or rights. Under maximizing theories law acts as a surrogate for these impracticable calculations. The calculation is made at an institutional level, by legislators or officials. Though it cannot in the nature of things be very precise, a vicarious calculation is better than self-assessment. It is relatively impartial. In making it legislators and officials act as moral agents whose task is to determine what will maximize certain values, including distributive justice. Indeed whenever impartiality is a moral requirement, it would seem that an institutional determination is called for. The determination may be made within families by parents and within other organizations by the appropriate authority, but at the political level only by state institutions, and in particular by law. How else could this aspect of morality be given effect? Of course it is true that when institutions make, or try to make, impartial determinations, they also take account of non-moral, instrumental reasons, like the need to maximize the GNP or to decide in a way consistent with authority. The moral remit is not an exclusively moral remit. So in fixing tax-rates account is also taken of ease of collection and in adjudicating on tax law of the principles of and precedents in this branch of the law.

In the upshot, calculation is not a viable determinant alternative to law. The third possible alternative is convention. The '*nomos*' in which, according to Aristotle, noble conduct or justice seems to consist can mean either convention or law. The difference is that conventions are not institutional but grow up informally by custom, usage, or practice, including trial and error, when people try to co-ordinate their activities. Convention and law are certainly closely related; and convention can often supply the necessary element of determination. Thus in a simple society there might be a convention that every household must supply an able-bodied man to serve in time of war. But in a more complex society it is clear that moral and political obligations cannot be fully determined by this means. Convention is not an effective means of settling problems that have not come up before, or not often enough to give rise to a settled practice. Nor does recourse to convention suffice when differences of opinion prevent the emergence of a settled practice, or when a minority flout the general practice. In that

[7] Could there, to take Raz's example in *The Morality of Freedom* (1986) 45ff, be a system of voluntary contributions to a state charity, with official guidelines for self-assessment, instead? I think his example is meant to show, and succeeds in showing, that the state in fixing taxes builds on a pre-existing (abstract) obligation, not that it would be a possible method of raising revenue.

case a law is needed. Moreover conventions need interpreting, and for moral reasons, interpretation must be impartial. So, at least at a political level, an institutional mechanism is called for by which some recognized authority is given power to interpret the convention. However the authority may be described, this constitutes a recourse to law, and the decisions of the authority form precedents. So while convention can partly fix the content of some moral obligations, it cannot provide for their interpretation. Particularly in a complex society, it cannot settle the content of others at all.

It may be objected that in the cases described law acts as an indirect method of establishing a convention, and that it is this convention, not the law, that forms part of morality. But many laws, like tax laws, do not work by seeking to establish conventions. Others, like traffic regulations, may aim to establish new and better practices, in this case driving practices. But often the desired behaviour could not become established without a law, and moreover a law which possessed moral force.

I have dealt with calculation and convention as determinants alternative to law. The fourth possibility is resort to commands. The moral and political importance of the distinction between laws and commands has perhaps not been fully grasped. The term is used here in the narrow sense of orders which are not open to legal or moral challenge. If they are open to challenge, like the orders of a policeman or an army officer according to contemporary western notions of law, they are subordinate aspects of law and not something distinct from it.

It would seem at first sight that unchallengeable commands might be the sole determinants of a morality which treated unquestioning obedience to superiors as a virtue of overriding importance. Perhaps no rational moral theory takes this form, but as some have thought otherwise, it is better not to cut short the discussion by too narrow a definition of co-operative morality. Given a moral theory, then, of this sort and a community with widely shared moral opinions, could their determinants take the form of commands issued by those in authority to give effect to prevailing moral opinion?

Here I insert a historical digression, which concerns the relations of law and morality at a traumatic period of recent history.[8] A command-oriented moral theory of the sort just mentioned dominated the Third Reich. The theory was, in bare outline, composed of two elements along with a connecting link. The test of right and wrong was taken to be sound popular feeling, the moral opinions of the masses. These opinions were supposed to draw a sharp distinction between members and non-members (or unworthy members) of the community, for instance Aryans and non-Aryans. As regards members the morality was co-operative, as regards non-members hostile. But this morality in both its aspects was given effect by means of a second principle, the Leader Principle, which laid down that the commands of superiors be obeyed without question, whatever their content. It was not for the recipient of a command to judge whether it

[8] On what follows see *Recht und Justiz im Dritten Reich* (ed R. Dreier and W. Sellert, Suhrkamp, Frankfurt 1989).

conformed to the test of sound popular feeling. These elements of Nazi morality were linked by ensuring that those in authority, entitled to issue binding commands, were committed to the morality of sound popular feeling.

This enabled Hitler, in principle at least, to substitute commands for laws as an instrument of government. Indeed he seldom resorted to legislation and was reluctant to embody in legislative form the regime's most radical policies. For example the compulsory euthanasia programme,[9] by which some seventy thousand people who were terminally ill or suffering from disabilities were put to death, was carried out on the basis of verbal orders. These ran counter to the Criminal Code, whose provisions as regards homicide remained in force throughout the period. But when this was pointed out Hitler rejected the suggestion that the programme should be put on a statutory basis, though in the upshot it was gradually run down. More strikingly still, the holocaust itself, the Final Solution, lacked any legal basis; it was launched and carried through by verbal orders passed down from the top. It too was a violation on a grand scale of the Criminal Code.

This example suggests that for a rather special type of moral theory commands, in the narrow sense of orders not open to moral or legal challenge, can serve as a determinant of behaviour to the exclusion of law. But it also suggests that resort to command as a substitute for law is most likely to occur when those who have power are ashamed openly to declare how they propose to give effect to the morality they espouse. It is resorted to when that morality is either unavowed or, if it is avowed, when the consequences to be drawn from it—its determinants—remain unavowed. The part of Nazi morality that was hidden or half-hidden, and put into effect through command rather than law, was that of hostility to those whom popular feeling allegedly treated as non-members of the community. This example shows that a morality of hostility can have commands as determinants. It does not show that a morality of co-operation can be made determinate by commands alone.

Admittedly this is not the only possible interpretation of these terrible events. In particular Hitler saw the matter differently. He said on several occasions that in the Third Reich there was no distinction between morality and law. By this he undoubtedly meant that the commands of those in authority, assumed to be issued on the basis of sound popular feeling, had the same force as the formal provisions of the legal system. Presumably, though this was not specified, later commands overrode earlier laws to the extent that the two were inconsistent. A rather similar analysis must account for the conclusion reached by a scholar with very different views. After the war Radbruch, equating laws and commands, wrote that the legal outlook of the Nazi regime was positivist, and that positivism in the Weimar republic had prepared the way for it.[10] Investigation has shown that view to be wrong.[11] Legal positivism did not prevail in either the Weimar

[9] M. Diesselhorst in *Recht und Justiz*, above n 8, 118–35.

[10] G. Radbruch SJZ 1946 no 5 105–8 where he speaks of the two (positivist) principles 'orders are orders' and 'statutes are statutes'.

[11] M. Walther in *Recht und Justiz*, above n 8, 323–34.

republic or the Third Reich, though Radbruch, himself both a positivist and a democrat, was an exception. The lawyers and especially the judges of the Weimar republic, hostile to democracy, demanded the right to test the validity of statutes, since they looked on the Weimar legislation as vitiated by democratic notions. Rejecting judicial neutrality, the German Judicial Association dissolved itself within a few months of Hitler's coming to power and joined the equivalent Nazi organization *en masse*. Building on this foundation the Nazis insisted that laws be interpreted in the sense required by sound popular feeling, so that, as Carl Schmitt put it, the legal system could be radically transformed without altering a single statute.[12] In practice judges were not as radical as all that. A more sober description of what happened in the Third Reich is that commands replaced law as regards relations between members of the community and their enemies, ie those who, on the sound popular feeling test, were non-members. But law was kept, largely in its existing form, as a determinant of the co-operative morality which the regime encouraged for Aryans among themselves.

It is only if commands are treated as laws that either Hitler's view or Radbruch's can appear plausible. Hart argued with force that they were not equivalent.[13] It seems to me that what is at stake is this. Commands can certainly be determinants of a morality of hostility. Indeed the substitution of command for law in the treatment of those regarded as enemies is a normal feature of oppressive regimes. How far commands can feature as determinants of a morality of co-operation is more debatable. Perhaps we should distinguish, within schemes of co-operation, between the rational and the non-rational. A morality which demands unquestioning obedience to commands is not a rational morality in the sense in which in the western tradition law is rational. It does not present those whom it addresses with genuine, though weighted, choices. It is a morality which works, and perhaps may succeed in ensuring co-operation, primarily (and not, as law does, secondarily) by instilling fear and foreclosing choice. If this is correct, then it is only in the context of a rational morality of co-operation that there is a sharp distinction between commands and laws. The determinants of such a rational morality must include laws.

One final complication as regards commands. In societies where the prevailing morality differentiates between groups, and where commands displace laws in the treatment of certain of them, always to their disadvantage, those who have authority to issue commands are usually designated by law. That was true in the Third Reich and equally in the other oppressive societies I can think of. But even in tolerant societies there are recognized differences between members and non-members. Non-members—for example aliens—are often subject to orders which they are unable effectively to challenge. That law is a necessary determinant of morality is perhaps true only of a morality of co-operation in the strict sense. Whether it is also true of a morality which regulates relations with strangers on a basis of co-existence or tolerance requires further consideration.

[12] C. Schmitt, *Ueber die drei Aufgaben der rechtwissenschaftlichen Denkens* (Hamburg 1934) 59.
[13] *The Concept of Law* (1961) 18–25.

It seems, then, that neither calculation, convention nor command can wholly displace law as the determinant of a rational co-operative morality. But does this apply to the whole of morality, or only to certain parts of it? The need for a determinant, we saw, does not depend on the particular moral theory adopted. But it has been assumed that the theory adopted calls for certain things to be made compulsory. Yet moral theories, if construed so as to include reflection about elevated ideals of conduct, do not yield only requirements. They also recommend ways of life and the pursuit of goals which are not treated as compulsory. Indeed not all moral ideals can be compulsory. Generosity is a virtue, but that we should be generous can be no more than a recommendation. Compulsory generosity would cease to be generosity. Perfectionism could not be compulsory for the same reason; it consists in going well beyond what is obligatory. For these and other voluntary aspects of morality a legal determination is neither needed nor possible. One may contrast with generosity, which must be voluntary, the virtue (call it a sense of social responsibility) which disposes people to contribute to the needs of their community. There could certainly be a law defining generosity, and perhaps laying down that a person who in terms of the definition counted as generous was entitled to a tax exemption. But the exercise of generosity does not depend on there being a legal determinant of what it is to be generous. On the contrary, the legal definition would not make a person who conformed to it generous if their conduct did not count as generous without it. It would have no moral force. On the other hand the legal determination of what a taxpayer owes by way of tax is necessary to the exercise of social responsibility, because social responsibility, unlike generosity, is a disposition to do, at a minimum, what is formally required of one. So a legal determination of social responsibility has a moral status, though not a conclusive one.

In the upshot the minimum parts of morality that call for determination are those which make conduct compulsory. The legal determinants of these need not however necessarily impose obligations or duties. Many laws are in no way normative.[14] They serve rather to define the conditions in which conduct is required or optional and to provide facilities for those who want to take advantage of the options open. For instance the duty to support one's family calls for determinants. But these need not relate directly to rights of support. They may include laws which confer on those who have a duty to support their families the power to make wills or set up trusts. Again the duty to protect others may call for legal determinants which include the power in certain conditions to punish wrongdoers. Moreover the need for determination varies from one sort of obligation to another. The duties of members of a community to that community can hardly be fixed except by law. That is less true of the obligations of members of a community, or of states in the international community, to one another. Here the difficulty is not, as with obligations owed to a community by its members, that no one's obligation can be determined unless everyone's is. On

[14] T. Honoré, *Making Law Bind* (1987) 69–88.

the contrary, abstract reasoning or convention will often indicate within limits what is required of each individual or state by way of, say, consideration for others. But complex problems of co-ordination cannot be solved in this way. For example respect for others and for efficiency requires that the use of wavelengths for transmitting information be co-ordinated. Interference both within and between states must be minimized. But no person or state has an obligation to refrain from using a particular wavelength unless wavelengths are allocated by municipal or international law. The same is true of other complex problems of co-ordination, for example in running transport, promoting health, or preserving the environment. In regard to these there is often no possibility of doing the right thing apart from law since there is, apart from law, no right thing to do.

It seems clear, then, that legal determinants of moral obligations are necessary to a co-operative morality whenever complex problems of co-ordination arise. Morality is not fully autonomous. Moreover its dependence fits the view, admittedly disputed, that when laws lay down or imply the existence of obligations they assume that these obligations have moral force. If laws serve as determinants of morality, it is easy to see why this assumption is made. That some laws do have moral force need not be just a self-interested claim on the part of governments or lawyers. It may be a necessary assumption for a society which aspires to having a viable morality. Of course to claim moral force is not the same thing as to possess it. But sometimes, if the legal determinant of a moral obligation does not have moral force, there will be a vacuum.

Legal Ethics: Professional Ethics and Personal Integrity

Conference
University of Auckland, New Zealand □23 - 25 June 2006

ZOË PREBBLE[1] AND JOHN PREBBLE[2]

THE MORALITY OF TAX AVOIDANCE:

WHY THE LEGAL DIFFERENCE BETWEEN EVASION AND AVOIDANCE IS INSUFFICIENT TO GROUND A MORAL DISTINCTION

[1] BA (Hons), LLB (Hons) Victoria University of Wellington, The New Zealand Law Commission, Wellington.
[2] BA, LLB (Hons) Auckland, BCL Oxon, JSD Cornell, Inner Temple, Professor and former Dean of Law, Victoria University of Wellington, Senior Fellow, Taxation Law and Policy Research Institute, Monash University, Melbourne, Henry Lang Fellow, Institute of Policy Studies, Wellington. The Lang Fellowship partially supported the work on this paper.

CONTENTS

I INTRODUCTION 4
II TAX EVASION AND TAX AVOIDANCE 4
A Defining Tax Avoidance 6
B Income-Splitting Schemes: Mangin v Commissioner of Inland Revenue 7
C Schemes for Deducting Inflated Expenses: Cecil Bros Pty Ltd v Federal Commissioner of Taxation 8
D Income Reducing Schemes: Jones v Garnett (Inspector of Taxes) 8
E Duke of Westminster's Case 9
F The Legal Line Between Tax Avoidance and Tax Evasion 9
G Blurriness of the Legal Line: Secrecy as a Condition for Successful Tax Avoidance 10
H Is Tax Avoidance Robustly or Only Contingently Legal? 10
I The Strict Construction of Tax Statutes 11
J Motive is not a Relevant Legal Consideration 11
K Assumptions about the Morality of Tax Avoidance 12
L Is the Moral Line the Same as the Legal Line? 13
M A Logical Error: The Slip From Legal to Moral Language 14
N The Alleged Morality of Tax Avoidance Requires Justification 14
O The Placement of the Moral Line: Two Alternatives 15
P The Spectrum of Tax Minimisation Behaviour: GAARs and Equivalent Doctrines 15
Q The Four Assumptions Behind the View that Tax Avoidance is Moral 16
III THE FIRST ASSUMPTION BEHIND THE VIEW THAT TAX AVOIDANCE IS MORAL: THAT THERE IS A MORAL ENTITLEMENT TO PRE-TAX INCOME 17
A Nagel's and Murphy's Response to the Assumption 17
B The Philosophical Underpinnings of the Assumption 18
C The Logical Incoherence of the Assumption 18
D Consequences of Rejecting the Assumption 19
IV THE SECOND ASSUMPTION BEHIND THE VIEW THAT TAX AVOIDANCE IS MORAL: THAT TAX AVOIDANCE AND TAX EVASION ARE NOT HARMFUL 20
A Culpability as a Moral Justification for Criminalizing Conduct 20
B Harmfulness as a Moral Justification for Criminalizing Conduct 21
C Diffuse Harms and Harms That are Difficult to Identify 21
D The Harmfulness of Tax Avoidance 23
E Kantian Categorical Imperatives 24
F Wrongfulness as a Moral Justification for Criminalizing Conduct 25
V THE THIRD ASSUMPTION BEHIND THE VIEW THAT TAX AVOIDANCE IS MORAL: THAT TAX EVASION IS MALUM PROHIBITUM 25
A Mala Prohibita and Mala in Se 26
B Mala Prohibita and Mala in Se: Mutually Exclusive Concepts 26
C The "Arbitrariness" of Tax Law 28
D Is It True that Tax Evasion is Only Malum Prohibitum? A Third Category of Wrongfulness 29
VI THE DEPENDENCE OF MORALITY ON LAW 30
A The Fourth Assumption: That all Aspects of Political Morality are Determinate and Independent of the Law 31

B Honoré: Some Aspects of Morality Depend on External Definition 32
C Primitive and Pre-Legal Societies 32
D Modern Legal Societies 33
E The Comparison of Primitive to Complex Societies. 34
F Tax Evasion and the Middle Ground Between Mala in Se and Mala Prohibita: A Legally Constructed Moral Wrong? 34
VII CONCLUSION 35

I INTRODUCTION

Tax evasion and avoidance involve conduct that is factually similar but legally distinct. Both aim to reduce or minimise tax liability, but avoidance is legal whereas evasion is illegal.[3] This paper responds to a line of judicial authority that stands for the general proposition that there is a moral entitlement to avoid taxes. The proposition has some intuitive and popular appeal but this paper will demonstrate that it is not supported by sound reasoning, but predicated on deeply flawed assumptions.

The conclusion that tax avoidance is moral appears to rest on four main assumptions that will be set out and critiqued in the course of this paper. Section III addresses the first assumption: that taxpayers have a moral entitlement to their pre-tax incomes. Section IV discusses a second assumption: that tax avoidance and evasion are not seriously harmful and therefore are not immoral. It will clarify the harms that both types of conduct cause. Section V will address a third key assumption: that tax evasion is malum prohibitum rather than malum in se. That is, the sole moral content of tax evasion is derived from its legal status. According to this assumption, since avoidance lacks illegality, the one quality that differentiates it from evasion and the one source of evasion's immorality, then avoidance must be moral. This section will demonstrate that despite the traditional conception of mala prohibita and mala in se as mutually exclusive and exhaustive categories, there is logical space for other hybrid types of legal wrongs to exist between the two extremes. The assumption to be addressed is that morality exists wholly independently of the law; this will be addressed in section VI with particular reference to the views of Professor Tony Honoré.

It will be demonstrated that not only do all four assumptions fail to justify the view that tax avoidance is moral, but that the evidence supports the opposite view.

II TAX EVASION AND TAX AVOIDANCE

Tax evasion and tax avoidance involve similar taxpayer behaviour and are each undertaken in pursuit of the same broad aim: to reduce or minimise tax liability. They are factually similar, but legally distinct.

[3] One of the present authors has argued elsewhere that, on examination, transactions that seem to involve no more than avoidance are more correctly categorized as evasion. John Prebble, "Criminal Law, Tax Evasion, Shams, and Tax Avoidance: Part i - Tax Evasion and General Doctrines of Criminal Law" (1996) 2 NZ Journal of Taxation Law and Policy, 3-16; "Part ii - Criminal Law Consequences of Categories of Evasion and Avoidance" ibid 59-74. See also sections IG and IH, below.

Tax evasion is illegal. It consists in the wilful violation or circumvention of applicable tax laws in order to minimize tax liability.[4] Tax evasion generally involves either deliberate under-reporting or non-reporting of receipts, or false claims to deductions. This conduct is legally straightforward to identify; a taxpayer has committed tax evasion only if he has breached a tax law. Indeed, evasion ordinarily involves criminal fraud.

Tax avoidance is not illegal. Rather, it is the act of taking advantage of legal opportunities to minimize one's tax liability.[5] Since a tax-avoider engages only in legal tax behaviour, even when the Commissioner of Inland Revenue detects his avoidance scheme, he is not ordinarily subject to criminal punishment. This is not to say however that the Commissioner will always allow legal tax avoidance schemes to stand.

Section BG 1 of the Income Tax Act 1994 is the most potent general anti-avoidance rule, known as a GAAR, that the Commissioner of Inland Revenue has at his disposal. It provides that:

> (1) A tax avoidance arrangement is void as against the Commissioner for income tax purposes.
> (2) The Commissioner, in accordance with Part G (Avoidance and Non-Market Transactions), may counteract a tax advantage obtained by a person from or under a tax avoidance arrangement.

Section OB 1 is the definitions section. "Arrangement" is defined broadly and includes any "contract, agreement, plan, or understanding (whether enforceable or unenforceable) including all steps and transactions by which it is carried into effect". The section also widely defines "tax avoidance" to include "relieving," "altering" and "avoiding, reducing or postponing" the incidence of tax.

Section GB 1 sets out the consequences that follow from an avoidance arrangement being void under section BG 1. Section GB 1 provides that the Commissioner may adjust the amounts of gross income, allowable deductions and net losses associated with the arrangement "as he thinks appropriate, so as to counteract any tax advantage obtained … under that arrangement". Essentially, in so adjusting these figures, the Commissioner may have regard to the business reality of the transactions that would have eventuated but for the arrangement.

[4] Bryan A Garner *Black's Law Dictionary* (8 ed, West Publishing Co, St Paul, Minnesota, 2004) 1501.
[5] Garner, above n 1, 1500.

Together with sections GB 1 and OB 1, section BG 1 allows the Commissioner to undo or ignore avoidance schemes that taxpayers may have devised. Suppose a tax avoidance scheme has the effect of splitting income or of generating deductions.[6] In such cases section GB 1 would operate to unwind the scheme's surface effect by "un-splitting" the income or "un-generating" the deductions. The GAAR allows the Commissioner to look behind legal appearances of tax avoidance schemes and to give effect to the true substance of the transactions.

A Defining Tax Avoidance

Unlike tax evasion, avoidance is not easily defined. While section OB 1 provides a definition of tax avoidance, this definition is too broad to be directly helpful. Taken literally, section OB 1 defines tax avoidance in such a manner that it includes many forms of tax activity that Parliament would have intended to let stand or even to encourage. For instance, if I make a donation to charity, I will qualify for a tax rebate; if I give income-producing assets to my children, I reduce my tax and I reduce the overall tax of the family as the income enjoys the lower tax rates of the children. In the language of section OB 1, both are examples where I have "altered" or "reduced" my incidence of tax. Yet everyone accepts that section BG 1 does not affect these arrangements. According to a literal interpretation of OB 1, these are "tax avoidance arrangements", but it is clear that Parliament did not intend that they should be void against the Commissioner of Inland Revenue. The rebate that I qualify for seems to be intended to encourage just this sort of donation, and no one would suggest that parents should not be able to make valuable gifts to their children.

The State often provides tax incentives of one kind or another. If taxpayers respond by investing in relevant industries and thereby achieve a reduction in their tax liability, then they have succeeded in "relieving", "altering" and "avoiding, reducing or postponing" their incidence of tax. Yet, this conduct is not contrary to the policy behind the tax incentives in question. Almost every case dealing with section BG 1, or an equivalent GAAR in another country, recognizes that GAARs are not intended to catch activities that Parliament seeks to encourage.[7] Courts occasionally use the expression "tax mitigation" to refer to measures that reduce tax (or avoid it, if one

[6] See *Mangin v CIR* [1971] NZLR 591 (PC); *Peterson v IRC* (2005) 22 NZTC 19 (PC).
[7] For example, *Challenge Corporation Ltd v Commissioner of Inland Revenue* [1986] 2 NZLR 513, 561 per Lord Templeman (PC).

will) but in circumstances such as those described in this and the previous paragraph: that is, in cases where the reduction of tax is the result of the taxpayer adopting a course of action that is clearly (and, ordinarily, expressly) encouraged by the relevant legislation.[8]

Nevertheless, a more precise definition of tax avoidance cannot be given. Clear examples of tax avoidance are available. They tend to exhibit such qualities as "artificiality", "undue complexity and circularity" or "lack of business reality".[9] However, avoidance itself defies comprehensive definition. We can begin to fill in a picture of avoidance through listing examples, but gaps remain. It seems we know tax avoidance when we see it, but we have to see it to know it.[10] It is useful here to consider a few examples of tax avoidance arrangements that have reasonably uncomplicated facts.

B *Income-Splitting Schemes:* Mangin v Commissioner of Inland Revenue

The case of *Mangin v Commissioner of Inland Revenue* [11] involved a tax avoidance scheme of a kind known as "income splitting". The taxpayer, a Canterbury farmer, owned six paddocks. The land was only fertile enough to produce crops every sixth year and had to be left to pasture for the rest of the time. The taxpayer rotated the cropping paddock every year. In any one year, the cropping paddock was the most profitable.

The taxpayer established a trust with his wife and children as beneficiaries. Each year he let the cropping paddock to the trust at a low rental. The trust engaged him as a contractor to sow the paddock in wheat and then to harvest it. The profits went to the trust. In respect of that paddock the taxpayer derived only rent and fees for his work. This rotation was repeated from year to year. The taxpayer thus reduced his income and incidence of tax. The trust was taxed each year for the income derived from the cropping paddock but at a lower tax rate than the taxpayer would have suffered.

---

[8] Idem.

[9] *Barclays Mercantile Business Finance Limited v Mawson (Inspector of Taxes)* [2005] STC 1, para 24 (HL), citing Park J in the High Court..

[10] Compare *Newton v Commissioner of Taxation* [1958] AC 450, 466, 98 CLR 1, 8 (PC), Lord Denning, judgment of the Board.

[11] *Mangin v CIR* above n 3.

The Privy Council agreed with the Court of Appeal's finding that the scheme fell within a general anti-avoidance provision and that as a result the scheme should be ignored for tax purposes.[12]

C *Schemes for Deducting Inflated Expenses:* **Cecil Bros Pty Ltd v Federal Commissioner of Taxation**

The taxpayer in *Cecil Bros Pty Ltd v Federal Commissioner of Taxation*[13] was a shoe retail company. It engaged in a tax avoidance arrangement that involved it paying higher than the market prices for some of its trading stock. Instead of buying from its usual supplier, the taxpayer bought some of its trading stock from a third company, Breckler Pty Ltd. Breckler bought the stock wholesale and then sold it to the taxpayer at increased prices. The taxpayer was willing to pay higher prices for the trading stock because related parties who enjoyed lower rates of tax owned Breckler.

The Australian High Court held that the taxpayer was entitled to deduct the extra costs that it had incurred through the scheme. It did not matter that there was an avoidance purpose; the costs related to the acquisition of trading stock and the economic benefit to Breckler was not legally relevant.

D *Income Reducing Schemes:* **Jones v Garnett (Inspector of Taxes)**

In *Jones v Garnett (Inspector of Taxes)*[14] the taxpayer, a specialist in information technology, set up a business with his wife. They established a company with two shares. The taxpayer and his wife each bought one share for a nominal sum. The taxpayer did not have a written employment contract but worked as director and employee of the company. The taxpayer represented all the earning power of the company, which provided his services to outside users for fees. His wife performed part-time administrative duties for which she received a fair salary. The taxpayer's salary was much lower than his experience and skills could command. The scheme enabled the company to earn profits, which were then distributed as dividends to the two shareholders, each being taxed at progressive rates. This resulted in lower overall tax liability for the couple than if the dividends had been paid solely to the taxpayer or if he had received a realistic salary and been taxed on that.

---

[12] *Mangin v CIR* above n 3, 598 Lord Donovan.
[13] *Cecil Bros Pty Ltd v FCT* (1964) 111 CLR 430.
[14] *Jones v Garnett (Inspector of Taxes)* [2005] All ER 396 (HC) Park J.

The existence of a specific anti-avoidance rule allowed the Revenue to unravel the scheme and to tax the dividends paid to the wife as part of the husband's income.

E Duke of Westminster's Case

In *Duke of Westminster v Commissioners of Inland Revenue*[15] the taxpayer had a number of employees, such as servants and gardeners, who were all on fixed wages or salaries. The taxpayer covenanted to make weekly payments to them of the same amounts they would otherwise have received as salary. The deed did not prevent the employees from collecting their ordinary salaries in addition to the covenanted payments. However, the employees wrote letters to the taxpayer in which they agreed that in practice they would forego their wage rights in consideration for the payments under the deed.

This arrangement qualified the taxpayer for a surtax deduction in respect of the deed payments. He would not have received a deduction if he had paid the employees' salaries in the usual way because the expenditure related to the Duke's personal needs, not to the earning of income. The arrangement was within the letter of the law but clearly was outside of its spirit. The policy behind the relevant legislation was that if part of a taxpayer's income never actually belongs to him because, as an annuity, it belongs to someone else, then he should not pay tax on it. However, in this case the annuities effectively functioned as salaries.

The House of Lords, however, upheld the taxpayer's scheme because there was no GAAR and because the Lords adopted a strict approach to statutory construction, favouring the arrangement's form over its substance.

F The Legal Line Between Tax Avoidance and Tax Evasion

"The law draws a line" between tax avoidance and tax evasion.[16] This line may be fine, but it is supposed to be crisp, such that any set of facts will fall "on one side of it or the other".[17] By definition, tax avoidance falls on the "safe side", whereas tax evasion is on the "wrong side", of the line.[18] In practice, however, the line can become blurred in a way that definition alone does not suggest.

---

[15] *Duke of Westminster v CIR* [1936] AC 1 (HL).
[16] *Bullen v Wisconsin* (1916) 240 US 625, 630 Holmes J.
[17] *Bullen v Wisconsin*, above, n 10, 630 Holmes J.
[18] *Bullen v Wisconsin*, above, n 10, 630 Holmes J.

G Blurriness of the Legal Line: Secrecy as a Condition for Successful Tax Avoidance

Suppose that the facts of *Mangin* were repeated today so that a taxpayer establishes a trust, rents out his cropping field on a rotational basis, and sows and harvests as a contractor. In one sense, this arrangement is legal and is merely avoidance: technically, the transactions have the effect that the taxpayer claims they do and the prima facie consequence is that he is liable for less tax.

However, if a taxpayer repeated the *Mangin* scheme today and was challenged by the Commissioner, there is no way that the arrangement would be allowed to stand for tax purposes; *Mangin* is the direct authority for striking it down. The taxpayer's only hope would be that the Commissioner would not become aware of the scheme. Secrecy is a necessary condition for the scheme's success.

Even some schemes not yet tested by the courts may depend on secrecy for their success. For instance, consider the *BNZ Investments Ltd v CIR*[19] case. This case involved several transactions, including one that was particularly circular. The taxpayer did not even attempt to argue that this most circular and artificial of the transactions was not avoidance for the purposes of the GAAR. Yet, if the Commissioner had never noticed this transaction, it would have had the legal effect of reducing the taxpayer's liability. In the case, the taxpayer won for other reasons, which are not relevant to the present point.

H Is Tax Avoidance Robustly or Only Contingently Legal?

If an avoidance scheme cannot achieve a reduction in tax liability without secrecy, it is hard to see it as legal in any robust sense. A scheme that will clearly be struck down if it should ever be challenged seems to be only weakly or contingently within the law.

Such an avoidance scheme does not seem robustly legal because usually we think of legality as involving a strong predictive aspect.[20] When I say, "it is illegal for me to commit murder," only part of what I mean is that if I kill someone else with the appropriate intent, and am found guilty in Court, then I have broken the law. Surely I also mean that even if I am never found out or challenged, still I have acted illegally.

---

[19] *BNZ Investments Ltd v CIR* (2000) 19 NZTC 15.

[20] Oliver W Holmes "The Path of Law" in Joel Feinberg and Jules Coleman (eds) *Philosophy of Law* (6 ed, Wadsworth, Belmont, 2000) (article 1st pub'd 1897) 174, 174.

This analogy should not be stretched too far; getting away with murder is of a quite different order to getting away with tax avoidance. An undetected murder is still not even weakly legal. However, recognition that much of tax avoidance is only legal in a weak sense is necessary to properly reflect the predictive aspect of law.

I The Strict Construction of Tax Statutes

There is a strong tradition of construing tax statutes literally.[21] That is, the Crown can levy tax only by bringing taxpayers within the letter of the law, and if it fails to do so, the taxpayer is free from tax even if his case appears to be captured by the spirit of the law.[22] The emphasis is on the legal structure, and legal effect, of the transactions that make up a purported tax avoidance scheme.

This strict interpretation approach is important for the success of tax avoidance schemes. We have seen that tax avoidance is not just the structuring of one's affairs to receive a legal tax advantage; it is doing so in ways inconsistent with the underlying policy or aims of the law.[23] A tax avoidance scheme can succeed then if the Court confines itself to the literal interpretation of the formal dimensions of the tax law; a scheme is likely to fail if the Court looks beyond the form of transactions to whether their underlying substance is within the law's spirit, not just its letter.[24]

J Motive is not a Relevant Legal Consideration

According to this strict approach to statutory construction, courts are to analyse the transactions in question with reference to the exact words of the statute. A corollary is that taxpayer motive is irrelevant when considering whether the taxpayer has remained on the safe side of the legal line.[25] The important question is whether a transaction satisfies the letter of the law; all we can learn from motive is whether a transaction coheres with the law's spirit.

[21] *Partington v A-G* (1869) LR 4 HL 100, 122 (HL) Lord Cairns; see also *Ormond Investment Co v Betts* (1928) 13 TC 400, 434 (HL) Lord Atkinson; *Duke of Westminster v CIR* above n 9, 19 Lord Tomlin; *St Aubyn and Others v A-G* [1952] AC 15, 32 (HL) Lord Simonds; *FCT v Westraders Pty Ltd* (1980) 144 CLR 55, 60 Barwick CJ. Also see generally Preston, above, n 15; Richard Lavoie "Subverting the Rule of Law: The Judiciary's Role in Fostering Unethical Behaviour" (2004) 75 U Colo L Rev 115, 183.

[22] *Partington v A-G* above n 15, 122 Lord Cairns.

[23] Doreen McBarnet "When Compliance is Not the Solution But the Problem: From Changes in Law to Changes in Attitude" (Working Paper 18, Centre for Tax System Integrity, Australian National University, 2001) 7.

[24] Tanina Rostain "Sheltering Lawyers: The Organized Tax Bar and the Tax Shelter Industry" (Research Paper 04/05-20, New York Law School, 2005) 9.

[25] *Vestey's Executors v CIR* (1949) 31 TC 1, 81; *Craven v White* [1988] 3 All ER 495, 518 (HL) Lord Oliver.

K Assumptions about the Morality of Tax Avoidance

A key assumption seems to underlie this traditionally strict approach to tax statutes. This assumption is that evasion may be immoral, and those who engage in illegal evasion may be "bad men", but legal tax avoidance is not a moral wrong.[26] Taxpayers are assumed not only to have the legal ability to avoid tax liability, but also a corresponding moral entitlement to do so.

This assumption is evident in numerous tax avoidance decisions. Lord Tomlin famously said in *Duke of Westminster v CIR*[27] that "every man is entitled if he can to order his affairs so that the tax attaching under the appropriate Acts is less than it otherwise would be."[28] He did not expressly say whether he was referring to moral, or only to legal, entitlement, but Lord Clyde was explicit on this point in *Ayrshire Pullman Motor Services and DM Ritchie v IRC.*[29] He considered that an arrangement was "neither better nor worse" for the reason that it had a tax avoidance purpose, adding that "[n]o man in this country is under the smallest obligation, moral or other, so to arrange his legal relations ... as to enable the Inland Revenue to put the largest shovel into his stores."[30]

Other cases have confirmed that there is "nothing illegal or immoral",[31] and "nothing wrong" about transactions with tax avoidance purposes.[32] Tax avoidance is not in the least "fraud[ulent]" but a basic taxpayer entitlement;[33] and a tax avoider "neither comes under liability nor incurs blame".[34] In *Gregory v Helvering*, Learned Hand J rejected the notion that there is "even a patriotic duty to increase one's taxes".[35] In another case he similarly stated:[36]

> There is nothing sinister in so arranging one's affairs as to keep taxes as low as possible. Everybody does so, rich or poor, and all do right; for nobody owes any public duty to pay more than the law demands. Taxes are enforced exactions, not voluntary contributions. To demand more in the name of morals is mere cant.

[26] *Ratzlaf v United States* (1994) 510 US 135, 140-41 Ginsberg J.
[27] *Duke of Westminster v CIR*, above, n 9.
[28] *Duke of Westminster v CIR*, above n 9, 19 Lord Tomlin.
[29] *Ayrshire Pullman Motor Services and DM Ritchie* v IRC (1929) 14 TC 754 (HL).
[30] *Ayrshire Pullman Motor Services and DM Ritchie* v IRC, above, n 23, 763-64 Lord Clyde.
[31] *A-G v Richmond* [1909] AC 466, 475 per Lord Atkinson (HL).
[32] *Jaques v FCT* (1923-24) 34 CLR 324, 362 Starke J.
[33] *CIR v G Angus & Co* (1889) 23 QB 579, 593 (CA) Esher MR.
[34] *CIR v Fisher's Executors* [1926] AC 395, 412 (HL) Lord Sumner.
[35] *Gregory v Helvering* (1934) 69 F 2d 809, 810 (2d Cir) Learned Hand J; affirmed *Gregory v Helvering* (1935) 293 US 465, 469, Sutherland J.
[36] *CIR v Newman* (1947) 159 F 2d 848, 850-51 (2d Cir) Learned Hand J dissenting.

This line of authority has been very influential.[37] Many tax lawyers suggest that the idea of ethical or moral considerations impinging on the tax-planning domain is simply absurd.[38] Some in the media have promoted the view that taxpayers have "a legal and moral right to work out how to pay as little tax as possible".[39] Lawyers and taxpayers who wish to engage in tax avoidance tend wholeheartedly to agree with judges like Lords Tomlin and Clyde and Justice Learned Hand, seeing tax avoidance "as being not only legal ... which it is by definition ... but [also] thoroughly respectable".[40]

L Is the Moral Line the Same as the Legal Line?

There is a wide variety of possible tax minimization behaviour that taxpayers might engage in. This spans a continuum from illegal tax evasion, to tax avoidance, where taxpayers reduce or alter their tax liability in a way that is contrary to the spirit of the legislation, and then to the reduction or alteration of tax liability in a way that is in line with Parliament's intent. It seems clear that there is a moral distinction between illegally evading tax on the one hand and making a donation to charity and receiving a deduction on the other. Few would dispute that acceptable tax mitigation such as charitable giving is moral and that tax evasion is immoral. Accordingly the moral line must be somewhere between these two extremes. Avoidance sits between acceptable tax mitigation and tax evasion. This means there are several possibilities as to where the moral line lies. The moral line might fall cleanly between avoidance and evasion or it might clearly separate avoidance from mitigation. The remaining possibility is that the line cuts less neatly through avoidance itself so that some avoidance is morally acceptable and some is not. The issue then is where exactly the moral line is drawn.

---

[37] See Geoffrey Lehmann, "The Income Tax Judgments of Sir Garfield Barwick: A Study in the Failure of the New Legalism" (1982 - 1983) 9 Monash U L Rev 115.

[38] A tax practitioner quoted in Rick Taylor "Rusty Pipes is Simply Rusty, Says Tax Practitioner" (11 July 1994) Tax Notes Today 135, 135.

[39] See George Monbiot "Comment and Analysis: Publish and Be Damned: The super-rich are fleecing us by avoiding taxes, and it should be a matter of public record" (28 September 2004, final ed) *The Guardian* London 21.

[40] For an extreme example, see Peter Clyne *How Not to Pay Any Taxes: A Handbook for Tax Rebels* (Professional Publications, Wellington, 1979) 44. Clyne was, admittedly, notorious in his time. See also Erich Kirchler, Boris Maciejovsky and Friedrich Schneider "Everyday Representations of Tax Avoidance, Tax Evasion and Tax Flight: Do Legal Differences Matter?" (Discussion Paper 187, Department of Economics, Humboldt University of Berlin, 2001).

The prevalent view discussed above is that tax evasion is immoral while tax avoidance is moral.[41] By definition, tax evasion is illegal and tax avoidance is legal at least in the weak sense discussed earlier.[42] According to this view then, the moral line between different types of tax minimization behaviour coincides with the legal line.

M A Logical Error: The Slip From Legal to Moral Language

While it is true by definition that tax avoidance is legal and evasion is illegal, it is a logical confusion to draw moral conclusions from legal facts. The question of whether particular conduct is moral is a matter of morality, not a matter of law. It is true that taxpayers are legally entitled, at least in a weak sense, to avoid tax. However, moral entitlement does not necessarily follow from legal entitlement. It is this logical error that many judges, lawyers and commentators make when they assert that there is "nothing illegal or immoral" about tax avoidance; they have good evidence for the first claim but give no persuasive justification for the second.[43]

N The Alleged Morality of Tax Avoidance Requires Justification

The cases have not provided any principled reason for us to believe that taxpayers are morally entitled to avoid tax. There have been judicial pronouncements on both sides of the debate on the morality of tax avoidance.[44] Some judges have noted that tax avoidance is legal but have not considered it to be moral.[45] Lord Denning famously remarked that tax avoidance "may be lawful, but it is not yet a virtue".[46] However, on the whole, arguments for the immorality of tax avoidance have not seemed to catch the imagination of commentators, lawyers or taxpayers to the same degree as statements asserting the opposite.[47]

This debate need not be "never-ending", however.[48] Statements that tax avoidance is not immoral may be quite attractive to some taxpayers and lawyers; such statements apparently let them off the moral hook so that they are left free to pursue

---

[41] See section II J.
[42] See section II H.
[43] *A-G v Richmond*, above, n 25, 475 Lord Atkinson.
[44] Anthony Mason "Where Now?" (1975) 49 ALJ 570, 574.
[45] See for example: *Latilla v IRC* [1943] AC 377, 381 (HL) Lord Simon; *Ensign Tankers (Leasing) Ltd v Stokes (Inspector of Taxes)* [1992] 1 AC 655, 675-76 (HL) Lord Templeman; 681 Lord Goff. See also Preston, above, n 15, 49; Monbiot, above, n 33; Rostain, above, n 18, 5; Joe Thorndike "Tax History – Civilization at a Discount: The Morality of Tax Avoidance" (2002) 95 Tax Notes 664; KD Deane "Law, Morality and Tax Evasion" (1984) 13 Anglo-Am LR 1, 7.
[46] *Re Weston's Settlement* [1968] 3 All ER 338, 342 (HL) Lord Denning.
[47] *Re Weston's Settlement* [1968] 3 All ER 338, 342 (HL) Lord Denning.
[48] Mason, above, n 38, 574.

wholeheartedly their own economic interests. This kind of attractiveness, however, tells us nothing about the truth of these statements. The moral status of tax avoidance merits closer scrutiny.

O The Placement of the Moral Line: Two Alternatives

There are two broad alternatives regarding the placement of the line that separates moral from immoral tax minimization activity. One possibility is that the authorities discussed above are correct and that the legal line between evasion and avoidance maps exactly onto, or indeed defines, the moral line. According to this view, legal tax avoidance is morally permissible and we are able to conclude that tax evasion is immoral simply because it is illegal.[49]

The alternative possibility is that the moral line is different from the legal line, that is, the legality of tax avoidance a matter separate from its morality. This view allows that it may be true that many or all instances of tax evasion are immoral and that many instances of tax avoidance are morally permissible, but holds that the bald statement that "taxpayers are morally entitled to avoid tax *because* it is not illegal" is not necessarily true.

P The Spectrum of Tax Minimisation Behaviour: GAARs and Equivalent Doctrines

Tax avoidance and evasion occupy different places on a spectrum of possible tax minimisation behaviour. At one end are transactions like gifts to charity or the disposal of income producing assets. As discussed, these are within the policy of the law and are not intended to be affected by a GAAR. The situation is similar in jurisdictions without a GAAR. There would be not question of disallowance of the above examples of tax mitigation under the substance over form rule in the United States of America[50] nor under the fiscal nullity rule in the United Kingdom.[51]

More serious and thus a little further along the scale are avoidance schemes that, although artificial, happen to succeed; for instance, *Cecil Bros Pty Ltd v Federal*

---

[49] Arthur Seldon "Avoision: The Moral Blurring of a Legal Distinction Without an Economic Difference" in AR Ilersic (ed) *Tax Avoision: The Economic, Legal and Moral Inter-relationships Between Avoidance and Evasion* (The Institute of Economic Affairs, London, 1979) 1, 3.

[50] *Gregory v Helvering* 69 F2d 809, 811 (2d Cir 1934), upheld, *Gregory*, 293 US 470 is said to have legitimized the substance over form approach, Jay A Soled, " Use of Judicial Doctrines in Resolving Transfer Tax Controversies" 42 Boston College L Rev 587, 591 (2001) collects a number of references.

[51] The doctrine originated in WT Ramsay v Inland Revenue Commissioners [1981] 1 All ER 865 HL.

Commissioner of Taxation[52] and *Peterson v IRC*.[53] More serious again is avoidance that might succeed but in fact does not because of a GAAR or equivalent judge-made doctrine. That is, the tax-avoider does not know beforehand that the arrangement's success depends entirely upon its secrecy. Further along the seriousness scale is tax avoidance that will certainly not work if challenged. Avoidance that can only succeed if not discovered is only contingently legal and is just short of tax evasion at the most serious end of the scale.

Q The Four Assumptions Behind the View that Tax Avoidance is Moral

As has been explained, there is much judicial authority for the proposition that the legal and moral lines between different modes of tax minimization behaviour are the same. However, this authority is not well reasoned. Rather, it is based on unsubstantiated assumption. The rest of this paper will test these assumptions. It will be demonstrated that not only do these assumptions fail to justify the view that tax avoidance is moral, but also that the contrary view is supported by both logic and evidence.

The first unsubstantiated assumption is that taxpayers are morally entitled to their pre-tax incomes and that taxation is an unjustified governmental incursion on individuals' private property rights. A second assumption is that tax evasion and avoidance are not especially harmful and so not immoral on that basis. A third assumption is that the crime of tax evasion is malum prohibitum rather than malum in se; that is, rather than concerning conduct that on any terms is wrong, the crime derives its only moral content from its legal status. The fourth and final assumption is that morality exists wholly independently of the law. Once these assumptions are rejected, it becomes clear that mala in se and mala prohibita are not exhaustive categories and tax evasion is not immoral simply to the degree that it represents a breach of the general duty to obey the law; it is immoral in a much a much deeper sense. Since tax avoidance is so factually similar to tax evasion, and evasion is immoral in a deep sense, then avoidance is also immoral; it is not rendered moral by a distinction from evasion that is strictly, and only, legal.

[52] *Cecil Bros Pty Ltd v FCT*, above, n 7.
[53] *Peterson v IRC*, above, n 3.

III THE FIRST ASSUMPTION BEHIND THE VIEW THAT TAX AVOIDANCE IS MORAL: THAT THERE IS A MORAL ENTITLEMENT TO PRE-TAX INCOME

The traditional approach to interpretation of tax legislation is to construe it strictly and literally.[54] Tax law is often likened to penal law in this respect; just as there cannot be punishment without law, so too there cannot be taxation unless "[t]he Crown ... [can] make out its right to the [tax]".[55] Taxpayers have the fundamental right "to arrange their affairs to minimize the taxes they must fairly pay".[56]

Implicit in this approach is the assumption that individuals have a prima facie moral and legal right to their pre-tax income; if an individual accumulates property through commercial and social transactions, the state cannot deprive him of it without unambiguous legal authority.[57] Private property rights are viewed as the natural, free-market ordering of things.[58] Taxation is thus a largely unjustified governmental interference in this natural order. Even the limited levels of taxation that are "justified" are coercive takings by the State.[59]

A Nagel's and Murphy's Response to the Assumption

The intuition that there might be a moral entitlement to one's pre-tax income is familiar to most taxpayers who have ever looked at their payslips and noticed the difference between the pre-tax and post-tax income figures.[60] However, Liam Murphy and Thomas Nagel challenge this view.[61] The assumption that pre-tax market allocations of resources are an appropriate moral starting point from which to assess the morality of taxation systems needs to be substantiated, not just asserted. Murphy and Nagel reject the philosophical underpinnings of the assumption on the ground that it is logically incoherent.

---

[54] See generally Preston, above, n 15.
[55] *CIR v G Angus & Co*, above, n 27, 593 (CA) Esher MR.
[56] David A Lifson (on behalf of Tax Division of the American Institute of the American Institute of Certified Public Accountants) "Revenue Provisions in President's Fiscal Year 2000 Budget: Hearing Before the Senate Finance Committee, 106th Congress" 1999.
[57] AP Simister and Winnie Chan "On Tax and Justice" (2003) 23 Oxford J Legal Stud 711, 714.
[58] Neil Brooks "The Responsibility of Judges in Interpreting Tax Legislation" in Graeme S Cooper (ed) *Tax Avoidance* and the Rule of Law (IBFD Publications, Amsterdam, 1997) 93, 97.
[59] Simister and Chan, above, n 49, 714.
[60] Kevin A Kordana and David H Tabachnic "Tax and the Philosopher's Stone" (2003) 89 Va L Rev 647, 650; Deane, above, n 39, 1.
[61] Liam Murphy and Thomas Nagel *The Myth of Ownership: Taxes and Justice* (Oxford University Press, New York, 2002) 35; see also Stephen Holmes and Cass R Sunstein *The Cost of Rights: Why Liberty Depends on Taxes* (WW Norton & Co, New York, 1999); Brooks, above, n 50, 97; Simister and Chan, above, n 49, 714; Kordana and Tabachnic, above, n 52, 649-650.

B *The Philosophical Underpinnings of the Assumption*

The assumption that the free-market ordering of property is morally privileged draws on the Lockean concept of natural property rights. Other philosophers of the early modern period had argued that property is not a natural kind but instead something that has been created by the State or has arisen out of society-wide convention.[62] Bentham later asserted that "property and law are born together and die together. Before the law there was no property; take away the law, all property ceases."[63]

John Locke had a different view; he thought that property could exist independently of particular politics or conventions.[64] According to Locke's account, resources from the natural world are reduced to the ownership of individuals through their labour.[65] Government or the State are not necessary for the creation of property. Instead, Government and the State arise after property, as a necessary mechanism for protecting natural rights.[66]

C *The Logical Incoherence of the Assumption*

Murphy and Nagel reject the Lockean-inspired view that property rights exist independently of special legal or political conventions and that we have a moral right to our pre-tax, market-derived income; they consider such a view, which they call "everyday libertarianism," wholly incoherent.[67]

In a pre-legal state of nature, such as envisaged by Thomas Hobbes, there would not be a "market" to yield "free-market outcomes" in anything like the sense that the everyday libertarian uses these terms.[68] The market depends on law. A working legal system is a necessary precondition for the existence of "money, banks, corporations, stock exchanges, patents, or a modern market economy ... the institutions that make possible the existence of almost all contemporary forms of income and wealth".[69] A

---

[62] Thomas Hobbes *De Cive: The English Version* (Howard Warrender (ed) Clarendon Press, Oxford, 1983) (1st ed 1647); David Hume *A Treatise of Human Nature* (LA Selby–Bigge and PH Nidditch (eds) Clarendon Press, Oxford, 1978) (1st ed 1739) 489.
[63] Jeremy Bentham, *The Theory of Legislation* (CK Ogden (ed) Kegan Paul, Trench, Trubner & Co, London, 1931) (1st ed 1802) 113.
[64] John Locke *Two Treatises of Government* (Peter Laslett (ed) Cambridge University Press, Cambridge 1988) (1st ed 1689).
[65] Jeremy Waldron "Property" in Edward N Zalta (ed) *The Stanford Encyclopedia of Philosophy (Fall 2004 Edition)* <http://plato.stanford.edu> last accessed 1 August 2005).
[66] Waldron, above n 57, s4.
[67] Murphy and Nagel, above, n 53, 74.
[68] Hobbes, above, note 54; Kordana and Tabachnic, above, n 52, 650.
[69] Murphy and Nagel, above, n 53, 32.

legal system cannot exist without Government; Government depends on taxation.[70] Thus property rights fundamentally depend on taxation. It is therefore "meaningless" to speak of a prima facie property right in one's pre-tax income.[71] The distribution of resources via the free market thus has no position of moral privilege compared with other distribution models.[72]

The assumption that there is a moral right to pre-tax income arises in the context of the debate between legal positivists, like Bentham, and natural law theorists like Locke, and more recently between exclusive and inclusive legal positivists, about the relation between law and morality. It is beyond the scope of this paper to do more than simply draw attention to this jurisprudential context.[73] It is worth noting however that while Murphy and Nagel's position is non-Lockean, it need not be construed as purely positivist.

D Consequences of Rejecting the Assumption

If there is not even a presumptive moral entitlement to pre-tax income, then the imposition of taxes by the Government is not interference in the natural or moral order; any particular scheme of taxation and Government spending is simply one among many possible distribution models.[74] This alone is not to say that any particular taxation system is moral or immoral. Murphy and Nagel argue that questions of tax policy comprise just a small part of much wider set of questions of political philosophy and conceptions of justice. We can only evaluate the justice of

---

[70] Murphy and Nagel, above, n 53, 36. See also Holmes and Sunstein, above, n 53, 59; Leo P Martinez "Taxes, Morals, and Legitimacy" (1994) BYU L Rev 521, 540-41.
[71] Murphy and Nagel, above, n 53, 36.
[72] Murphy and Nagel, above, n 53, 59.
[73] See generally Leslie Green "Legal Positivism" *The Stanford Encyclopedia of Philosophy (Spring 2003 Edition)* Edward N Zalta (ed) <http://plato.stanford.edu> (last accessed 1 August 2005); Kenneth Einar Himma "Legal Positivism" *The Internet Encyclopedia of Philosophy* <http://www.iep.utm.edu> (last accessed 1 August 2005); Kenneth Einar Himma, "Natural Law" *The Internet Enclyclopedia of Philosophy* <http://www.iep.utm.edu> (last accessed 1 August 2005); Brian Leiter "Naturalism in Legal Philosophy" *The Stanford Encyclopedia of Philosophy (Fall 2002 Edition)* Edward N Zalta (ed) <http://plato.stanford.edu>; Mark Murphy "The Natural Law Tradition in Ethics" *The Stanford Encyclopedia of Philosophy (Winter 2002 Edition)* Edward N Zalta (ed) <http://plato.stanford.edu>; Feinberg, Joel and Jules Coleman (eds) *Philosophy of Law* (6 ed, Wadsworth, Belmont, 2000); HLA Hart "Positivism and the Separation of Law and Morals" (1958) 71 Harv L Rev 593; HLA Hart "Kelsen Visited" (1963) 10 UCLA L Rev 709; HLA Hart *The Concept of Law* (Oxford University Press, London, 1961).
[74] Murphy and Nagel, above, n 53, 36.

after-tax income by reference to the legitimacy of the wider political and social system in which it is set.[75]

Rejecting the assumption that there is a moral right to pre-tax income does not on its own allow us to conclude that tax avoidance is necessarily immoral. However, rejecting this assumption removes part of the basis for the conclusion that tax avoidance is moral.

IV THE SECOND ASSUMPTION BEHIND THE VIEW THAT TAX AVOIDANCE IS MORAL: THAT TAX AVOIDANCE AND TAX EVASION ARE NOT HARMFUL

In the context of criminal law, three key elements are often used to justify criminalizing particular conduct: culpability, social harmfulness and wrongfulness.[76] One way of analysing judges' statements that that tax avoidance is morally permissible is to consider tax avoidance, as compared with tax evasion, in the light of one of these three criteria for criminalizing conduct. These three criteria are not mutually exclusive; there is considerable scope for overlap between them. Nor are they exhaustive; there could be other reasons why a community might feel criminalization is justified. However, within the context of this paper, these criteria will provide a sufficient overview of the elements from which tax evasion and tax avoidance might derive any moral content.

A Culpability as a Moral Justification for Criminalizing Conduct

Tax evasion and avoidance each seem to satisfy the culpability element. Both share the same sorts of causes and motivations.[77] The tax avoiders and tax evaders alike seek to reduce or to avoid their tax liabilities. If anything, tax avoidance might often comprise a more involved and substantial mental element. The detailed planning of a tax avoidance scheme suggests a mind deeply engaged in the enterprise of minimizing taxes. There is some overlap here with the other two criteria for criminalizing conduct; if conduct is not harmful or wrongful, then its deliberateness need not mean that it is particularly culpable. But assuming that some harmfulness and wrongfulness

[75] Murphy and Nagel, above, n 53, 33; see also Michael Schler "Ten More Truths About Tax Shelters: The Problem, Possible Solutions, and a Reply to Professor Weisbach" (2002) 55 Tax L Rev 325, 395.
[76] See Stuart P Green "Why It's a Crime to Tear the Tag Off a Mattress: Overcriminalization and the Moral Content of Regulatory Offences" (1997) 46 Emory L J 1533, 1547.
[77] Anthony Christopher "The Law is the Law is the Law..." in AR Ilersic (ed) *Tax Avoision: The Economic, Legal and Moral Inter-relationships Between Avoidance and Evasion* (The Institute of Economic Affairs, London, 1979) 75, 80.

can be established, tax evasion and avoidance both seem to satisfy the culpability requirement.

B Harmfulness as a Moral Justification for Criminalizing Conduct

The harmfulness criterion is concerned with the quality of the act and its effects rather than with the characteristics of a particular defendant.[78] People and communities have interests, that is, those things that they have some stake in. An act is harmful to the extent that it intrudes on such interests.[79] Harmfulness is not a sufficient condition for criminalizing conduct; individuals' interests will often have some negative effect on each other but not all resulting intrusions will be harmful enough to justify criminal sanctions. Some conflicts, such as the potential conflicts of interest between competitors in a competitive marketplace, are seen as acceptable. However, harmfulness seems to be a necessary, or at least a close to necessary, condition; many commentators doubt whether conduct that is not harmful should ever be criminalized.[80]

The assertion that tax avoidance is not morally wrong perhaps relies to a certain extent on an assumption that it is not really very harmful conduct. The harm criterion can be difficult to assess with regard to so-called victimless crimes where it is difficult or impossible to identify a direct victim. However, tax evasion and avoidance do not fit within the definition of victimless crimes or victimless conduct; crimes tend to be considered victimless if only consenting adults are involved, for example, illicit drug use or consensual incest or consensual sexual deviancy.[81] Attempts to claim that avoidance is harmless because it is victimless cannot succeed.

C Diffuse Harms and Harms That are Difficult to Identify

Tax evasion is not a victimless crime in the sense described in the foregoing paragraphs but there is a superficially analogous argument that seems to be implicitly relied on to show that tax evasion is not harmful and so is morally neutral.[82] This argument is not that tax evasion is victimless in the standard sense of the term.

---

[78] Green "Why It's a Crime to Tear the Tag Off a Mattress" above, n 68, 1549-50.
[79] Green "Why It's a Crime to Tear the Tag Off a Mattress" above, n 68, 1550.
[80] See, eg, Andrew von Hirsch "Injury and Exasperation: an Examination of Harm to Others and Offense to Others" (Review) 84 Michigan L Rev 700 (1986).
[81] Garner, above n 1, 400.
[82] Barry Bracewell-Milnes "Is Tax Avoidance/Evasion a Burden on Other Taxpayers?" in AR Ilersic (ed) *Tax Avoision: The Economic, Legal and Moral Inter-relationships Between Avoidance and Evasion* (The Institute of Economic Affairs, London, 1979) 105, 107.

Instead, the argument begins by posing the focus question, "who are these victims and what is the exact harm?" The idea seems to be that it is very difficult to pinpoint any particular victim of tax avoidance or evasion. The argument proceeds from lack of identifiable victims to the conclusion that no harm is caused by tax avoidance and evasion.

The flaw in this argument is that it assumes that a lack of individually identifiable victims is the same thing as a lack of victims altogether and that sufficiently diffuse harm is substantially similar to a total absence of harm. It is true that if taxpayer X chooses to pay less than his lawfully level of tax, we cannot point to some other individual Y who is a direct victim of this conduct. But this does not mean that no one is affected by taxpayer X's avoidance or evasion of his taxes. He has caused harm, and this harm affects real people, but the harms are diffuse. These diffuse harms associated with tax avoidance and evasion are more difficult to visualise than the harms that affect the very tangible victims of street crime.[83] However, the harm criterion will be satisfied whether the harm in question is spread thinly but widely or concentrated on one identifiable victim. The harm caused by tax evasion may be significant only in the aggregate, and many of the victims of tax evasion may remain unaware of the harm that they suffer, but the harm is significant none the less.[84]

The problem with suggesting that diluted harms at some point cease to be real harms can be illustrated by reference to the dilemma of large numbers, a form of the prisoners' dilemma.[85] In a community, any individual knows that his decisions will affect others and that in turn their decisions may affect him. When individuals act within a group as large as a political community or society, the "others" become impersonalized and there is no way for any individual to control or reliably to influence the behaviour of all of his fellows.[86] If others seem to be complying with social duties, such as the duty not to freeload, or the duty not to harm others, then from a self-interested, game-theoretical point of view, there is little incentive for an

---

[83] Stuart P Green "Moral Ambiguity in White Collar Criminal Law" (2004) 18 Notre Dame J L Ethics & Pub Pol'y 501, 510.

[84] Green "Moral Ambiguity in White Collar Criminal Law" above, n 74, 509.

[85] Steven Kuhn "Prisoner's Dilemma" *The Stanford Encyclopedia of Philosophy (Fall 2003 Edition)* Edward N Zalta (ed) <http://plato.stanford.edu> (last accessed 1 August 2005).

[86] Jørn Henrik Petersen "The Moral Foundation of the Welfare State Versus the Mechanism: A Contribution to the Discussion of Philosophical and Theoretical Issues in Social Security Today" (Odense University, Denmark) <http://www.issa.int/engl/publ/2contjeru.htm> (last accessed 1 August 2005) 12.

individual to comply either.[87] The dilemma is this: for any individual it may appear that the harm that will result from refusing to comply with a social duty is so diluted as to be negligible; yet, if everyone followed that line and refused to comply, the negligible harms would add up to a very great harm. Eventually the system "bursts;" this is in no one's best interest.[88]

D The Harmfulness of Tax Avoidance

Economically speaking, there is no distinction between tax avoidance and tax evasion.[89] They are motivated by the same desire to minimize tax liability and have the same economic consequences. It is very difficult to gauge the amount of revenue lost to tax avoidance schemes, but even according to conservative estimates the sum lost in the United Kingdom each year to tax avoidance runs to tens of billions of pounds every year.[90] The United States tax avoidance boom of the 1990s was estimated to cost the federal government billions of dollars in lost tax revenue.[91]

Tax avoidance is harmful in that it results in a misallocation of resources.[92] Taxpayers spend time and money devising tax avoidance schemes. This expenditure of effort may be profitable for the taxpayer but only because he manages to extract an unintended tax benefit.

Tax avoidance therefore represents a deadweight loss to the economy as the taxpayer in achieving the tax benefit undertakes no actually beneficial activity.[93] The activity is demonstrably non-beneficial according to the following thought experiment: imagine that everyone, rather than just a subset of taxpayers, actively and aggressively pursued tax avoidance schemes wherever there was an opportunity to do so and there was little chance of being found out. The effect would be that tax rates would have to be raised and no one would achieve any gain. In fact, everyone would be worse off because tax avoidance is itself a deadweight cost.[94]

---

[87] Petersen, above, n 77, 12.
[88] Petersen, above, n 77, 14.
[89] Seldon, above, n 43, 3.
[90] Tax Justice Network figures quoted in Monbiot, above, n 33.
[91] Rostain, above, n 18, 2.
[92] Brooks, above, n 50, 96.
[93] Brooks, above, n 50, 96.
[94] O'Grady, above, n 83, 3.

Tax avoidance also undermines governments' revenues and their progressivity policies.[95] In practice, it has substantially negative distributional consequences.[96] Not all taxpayers are able or willing to devise or to take advantage of tax avoidance schemes.[97] Generally it is wealthy taxpayers or those with more sophisticated knowledge of tax law who are in the position to take advantage of tax avoidance opportunities.

Both avoidance and evasion risk undermining public confidence in the tax system. This can give rise to a vicious circle: as confidence falls, members of the public become less likely voluntarily to comply with tax laws.[98]

As a generalization, tax evasion is more likely to be carried out on a smaller scale than tax avoidance. For instance an individual taxpayer who engages in evasion by not declaring under the table income from his second job might in so doing deprive the Revenue of a few hundred dollars over a year. However, a company that structures a series of artificial transactions may generate legal tax deductions worth millions of dollars.

E Kantian Categorical Imperatives

Tax avoidance, and tax evasion for that matter, fail to satisfy the moral theory that Immanuel Kant called the "categorical imperative". The categorical imperative is a moral requirement deduced as a rational argument. Its essential postulate is that conduct is only moral if it would be acceptable even if all adopted it.[99] The argument from Kant's categorical imperative would proceed similarly to the point made above that tax avoidance is a deadweight loss to the economy and that for individual taxpayers to achieve gains through avoidance it is necessary that only a minority of taxpayers should engage in this behaviour. The categorical imperative would proceed

---

[95] Jeffrey Waincymer "The Australian Tax Avoidance Experience and Responses: A Critical Review" in Graeme S Cooper (ed) *Tax Avoidance and the Rule of Law* (IBFD Publications, Amsterdam, 1997) 247, 257; Doreen McBarnet "Law, Policy, and Legal Avoidance: Can Law Effectively Implement Egalitarian Policies?" (1998) 15 J L & Soc 113, 114.

[96] Brooks, above, n 50, 96.

[97] Brooks, above, n 50, 96.

[98] Michael O'Grady "Acceptable Limits of Tax Planning: A Revenue Perspective" (KPMG Tax Conference, Killashee, Ireland, 7 November 2003) 2.

[99] Graham Bird "Kant, Immanuel" in Ted Honderich (ed) The Oxford Companion to Philosophy (Oxford University Press, Oxford, 1995) 435; see also Robert Johnson, "Kant's Moral Philosophy" *The Stanford Encyclopedia of Philosophy (Spring 2004 Edition)* Edward N Zalta (ed) <http://plato.stanford> (last accessed 3 August 2005).

as follows: to see whether tax avoidance is moral, one must consider whether it would be rational behaviour for everyone to engaged in it. We have seen that universal tax avoidance cannot be rationally justified. Everyone would be worse off if all taxpayers avoided tax: tax rates would rise and since avoidance is a deadweight cost, all would be worse off. Tax avoidance thus fails to satisfy the categorical imperative.

F Wrongfulness as a Moral Justification for Criminalizing Conduct

In order to examine tax avoidance from an ethical perspective, this paper has employed an analogy with the three qualities that are ordinarily required of an action before it is criminalized, namely culpability, social harmfulness and wrongfulness. Foregoing sections of the paper have argued that tax avoidance satisfies the culpability and harmfulness requirements at least as well as does tax evasion. It follows that the only possible point of difference between the two is in terms of their respective wrongfulness. As a quality, wrongfulness focuses on the act rather than on the actor. A wrongful act is one that violates some kind of norm or moral standard.[100] However, wrongfulness is the very question at issue here; so consideration of this criterion cannot tell us much at this stage. The paper will revisit this factor in section VI.

V THE THIRD ASSUMPTION BEHIND THE VIEW THAT TAX AVOIDANCE IS MORAL: THAT TAX EVASION IS MALUM PROHIBITUM

An analytical framework that appears to have escaped earlier commentators is that tax avoidance and evasion, while legally distinct, are factually very similar. That is, the difference between evasion and avoidance is essentially a matter of law, not of relevant fact. Legal distinctions certainly provide sound bases for legal conclusions, but they cannot similarly justify moral conclusions. This observation is similar to the Humean point that one cannot derive an "ought" from an "is."[101]

To defend tax avoidance as legal and thus moral is not just to say something about tax avoidance but implicitly also is to comment on evasion. People who aver that there is a moral entitlement to avoid tax often make a point of contrasting avoidance

---

[100] Green "Why It's a Crime to Tear the Tag Off a Mattress" above n 68, 1551; Peter Aranella "Convicting the Morally Blameless: Reassessing the Relationship Between Legal and Moral Accountability" (1992) 39 UCLA L Rev 1511, 1530.

[101] Rachel Cohon "Hume's Moral Philosophy" *The Stanford Encyclopedia of Philosophy (Winter 2004 Edition)* Edward N Zalta (ed) <http://plato.stanford.edu> (last accessed 3 August 2005) s5.

and evasion; saying that there is a moral entitlement to avoid tax but no corresponding entitlement to *evade* it. This argument is hard to sustain when one takes into account that, the ingredients and effect of avoidance and evasion being factually similar, they are divided from one another by a line drawn according to law, not according to the facts of the case.[102] In contrast, generally speaking we expect that if two actions are factually similar, then they will be morally similar as well. There must be some basis for morally distinguishing two factually similar actions. If the only difference between evasion and avoidance is one of legality, and if avoidance is wholly moral then the immorality of evasion must be entirely attributable to its illegality. Put another way, it is not the content of the conduct itself that makes judges think that tax evasion is immoral, but simply that the evader has breached an overriding moral obligation to obey society's laws. Thus, judges return to the comparison between tax avoidance and evasion. Tax avoidance lacks the one characteristic that renders evasion immoral. People conclude that since avoidance is legal it must also be moral.

A Mala Prohibita and Mala in Se

There is a longstanding conceptual distinction between acts that are mala in se and those that are mala prohibita. An act that is malum in se is an evil in itself. For instance, murder, theft and rape are mala in se; such conduct is immoral because of its inherent nature, not because of its legal status.[103] Murder or rape would still be immoral even if not criminalized.[104] On the other hand, a crime that is malum prohibitum is a prohibited evil. For instance, jaywalking or not carrying a driver's licence while driving are mala prohibita.[105] Such conduct is criminal simply because it is prohibited by statute; it is not necessarily immoral in its own right. Mala prohibita and mala in se have customarily been understood as mutually exclusive terms:[106] acts are either wrong in their own right or only wrong because they have been prohibited.

B Mala Prohibita and Mala in Se: Mutually Exclusive Concepts

Although judges who identify tax avoidance as moral have not expressly used the terms malum in se and malum prohibitum, these concepts seem to underlie their

---

[102] Sanford H Kadish "Some Observations on the Use of Criminal Sanctions in Enforcing Economic Regulations" (1963) 30 Chi L Rev 424, 425.
[103] Garner, above n 1, 978.
[104] Stuart Green "Why It's a Crime to Tear the Tag Off a Mattress" above n 68, 1571; see also Patrick Devlin *The Enforcement of Morals* (Oxford University Press, London, 1968) 33.
[105] Garner, above, n 1, 978-79.
[106] Garner, above, n 1, 979.

views on the respective moral statuses of tax avoidance and tax evasion.[107] Since the terms are supposed to be mutually exclusive and exhaustive, any wrong that would not be a wrong independently of its illegality is by definition not malum in se but is necessarily malum prohibitum.

It is at least technically or trivially true that tax evasion is not malum in se. Tax evasion is the wilful attempt to defeat or circumvent the tax law so as to reduce one's tax liability.[108] If there is no tax law, there is nothing wilfully to defeat; that is, no conduct that could even arguably be independently immoral. For instance, unlike many jurisdictions, New Zealand does not have a capital gains tax. While there are circumstantial or policy reasons for the presence or absence of capital gains tax in any country, from the point of view of an individual taxpayer it is fairly arbitrary whether such a tax exists. Consider two identical taxpayers, each with the same business and earnings structures. One is based in the United Kingdom and the other in New Zealand. Each makes identical capital gains, but only the first taxpayer is liable to capital gains tax. We know that the New Zealand taxpayer is not legally guilty of capital gains tax evasion because of the absence of any law imposing that tax. But more than this, we are sure that he has done nothing that is morally wrong.

The judges who assert a moral entitlement to avoid tax seem to assume that it follows from the above observations that the United Kingdom taxpayer who evades the capital gains tax is morally culpable for breaking a valid law, but would not be morally culpable had that law not existed. Accordingly, if tax evasion is not malum in se, it must be malum prohibitum. If it is malum prohibitum, it is only immoral by virtue of its illegality. Thus, tax avoidance, which is not illegal, is wholly morally permissible.

It is true that evasion involves lying when it concerns the under-declaration of income. Evasion is very often achieved through criminal fraud, such as the fabrication of false accounts. It seems quite reasonable that conduct involving lies or fraud might be malum in se. But this does not mean that tax evasion itself is malum in se. The frauds or lies perpetrated by the tax evader may be highly wrongful independently of their legal statuses. Fraud and lies may accompany the non-payment, or under-payment of tax. But it need not follow that the tax evasion itself is malum in se.

---

[107] See also Roberta Romano *The Advantage of Competitive Federalism for Securities Regulation* (American Enterprise Institute, Washington, 2002) 60.
[108] Garner, above, n 1, 1501.

In practice tax evaders tend to embroider their transactions a great deal in order to generate deductible losses for instance. But tax evasion does not in theory require elaborate embellishment and often has not embellishment in practice. Perhaps the simplest form of tax evasion is the failure ever to file a return. Assuming that one has derived some income, the complete failure to file a return is certainly evasion since it is a deliberate attempt illegally to defeat the tax law. This may be a more austere kind of tax evasion than we are used to, but it is evasion nonetheless. Such evasion does not involve any lying, nor does it constitute fraud, since both categories seem to require more than a deliberate omission, that is, some positive act or statement.[109] So, this purest form of evasion is not accompanied by any actual lies or fraud. If it is morally wrong, it is not wrong in virtue of its association with other morally wrong behaviour.

C The "Arbitrariness" of Tax Law

Judges who say that tax avoidance is moral do not think that tax evasion is excluded from being malum in se in only a trivial or technical sense. They further suggest that even where there is a tax law relating to the conduct in question, the conduct's moral content does not extend beyond that which it derives purely from its illegality. They draw on tax law's perceived arbitrariness: tax law is arbitrary and an arbitrary law surely cannot be malum in se. Their persistent reference to line drawing in tax law seems calculated to suggest this arbitrariness.[110] If tax is an area governed by no universal moral imperative, then a legal line, wherever it is drawn, has no necessary relation to the facts that it governs.[111] Legal rules such as the rules of taxation are not logically deduced, but posited: "the claim of our [tax] code to especial respect is simply that it exists, that it is the one to which we have become accustomed, and not that it represents an eternal principle."[112]

[109] Garner, above, n 1, 685.
[110] *Bullen v Wisconsin*, above, n 10, 630 Holmes J.
[111] See Lord Oliver "Judicial Approaches to Revenue Law" in Malcolm Gammie and A Shipwrite (eds) *Striking the Balance: Tax Administration, Enforcement and Compliance in the 1990s* (Institute for Fiscal Studies, London, 1996) 174.
[112] Oliver W Holmes "Law in Science and Science in Law" (1899) 12 Harv L Rev 443, 460; see also Andrew Schotter *The Economic Theory of Social Institutions* (Cambridge University Press, Cambridge, 1981) 21.

D Is It True that Tax Evasion is Only Malum Prohibitum? A Third Category of Wrongfulness

We have seen that tax evasion logically cannot be malum in se; any talk of evasion of tax law where there is no tax law is meaningless. However, it is dangerous to proceed from that observation to the conclusion that tax evasion is necessarily malum prohibitum and that the conduct itself is therefore morally neutral.

The division of all wrongs into mala in se and mala prohibita has a long history. It is appealing because it seems to capture some observable truths: some wrongs like murder and rape do seem to be immoral independently of their legal categorization; some crimes do seem largely regulatory, such as jaywalking or failing to carry one's driver's licence when driving. However, the flaw that seems to have accompanied the terms mala prohibita and mala in se throughout their history is the notion that they are mutually exclusive and exhaustive. Most particularly, it is not self-evident that the categories are exhaustive of the possible relationships between the legality and morality of various types of acts.

Tax evasion may not be malum in se; however, it seems a quite different type of conduct from paradigmatic mala prohibita.[113] Even if tax evasion is not mala in se, there seems to be more wrong with it than is wrong with purely regulatory violations.[114] Tax crimes are generally classed as "regulatory" as they are governed by statutes that are administered by an administrative agency, that is, the Inland Revenue Department.[115] But mere definitional sleight of hand is not enough to justify the alleged moral entitlement to avoid tax. The very questions at issue are first, whether there is a moral duty not to evade tax and secondly, whether there is a similar duty not to avoid it. Blindly assuming that mala in se and mala prohibita are exhaustive of the moral landscape does not get us very far in answering these questions.

It is similarly circular to identify tax evasion as a "regulatory offence" and then to assume that all regulatory offences are mala prohibita such that their content is morally neutral.[116] Concepts like malum prohibitum and "regulatory offence" seem often to be called upon to do more explanatory work than they are logically capable of. The mere fact that a statute is administered by an administrative agency is no necessary indicator that the statute lacks moral content. The concepts malum in se and

[113] Garner, above, n 1, 978-79.
[114] Devlin, above, n 95, 16.
[115] Green "Why It's a Crime to Tear the Tag Off a Mattress" above n 68, 1544.
[116] Aranella, above, n 91, 1530.

malum prohibitum may have the weight of tradition on their side, but any proposition that they are exhaustive is based on assumption, not on reason. People who say that tax avoidance is not immoral seem to rely on a false dichotomy: it is not correct to say that unless a wrong is immoral entirely independently of all law, its content must be morally neutral and that its sole claim to moral weight must be derived from a general obligation to obey the law. There is plenty of logical space between these two paradigms for other hybrid types of moral relationship between law and moral obligation. The final part of this paper discusses this logical space.

VI THE DEPENDENCE OF MORALITY ON LAW

Most commentators and judges who assert a legal and moral right to avoid taxes would not go so far as to say there is also no moral duty not to evade taxes. That is, they are not sceptics about the existence of at least a prima facie moral duty to obey the law.[117] In their legal decisions, judges do not usually engage in detailed theoretical discussions of political obligation. The writers have found no tax decisions in which a judge has discussed the relative merits of social contract theories of political obligation as compared to fairness theories or any other model. It is a rare judge who would deny any prima facie obligation to obey valid law. It is beyond the scope of this paper to analyse all the possible derivations of such a duty to obey the law.[118] For the purposes of this paper, it will be sufficient to proceed on the same assumption as the judges, that is, that there is a general moral obligation to obey the law.

Judges who assert a moral entitlement to avoid tax derive a specific duty to obey tax law in particular from a general obligation to obey the law. They agree that since tax evasion is illegal there is a moral duty not to engage in it.[119]

---

[117] For some sceptical perspectives see MBE Smith "Is There a Prima Facie Obligation to Obey the Law?" (1973) 82 Yale L J 950; Richard Wasserstrom "The Obligation to Obey the Law" (1963) 10 UCLA L Rev 780; AD Woozley *Law and Obedience: The Arguments of Plato's Crito* (University of North Carolina Press, Chapel Hill, 1979); Joseph Raz "The Obligation to Obey the Law" in *The Authority of Law: Essays on Law and Morality* (Oxford University Press, Oxford, 1979) 233; Joseph Raz "Respect for Law" in *The Authority of Law: Essays on Law and Morality* (Oxford University Press, Oxford, 1979) 250; A John Simmons *Moral Principles and Political Obligations* (Princeton University Press, New Jersey, 1979).
[118] See generally A John Simmons, "Obligations, Political" in Edward Craig (ed) *Routledge Encyclopedia of Philosophy* (Routledge, London, 1998) vol 7, 76.
[119] Tony Honoré "Must We Obey? Necessity as a Ground of Obligation" (1981) 67 Va L Rev 39, 48.

A The Fourth Assumption: That all Aspects of Political Morality are Determinate and Independent of the Law

The picture of tax evasion as malum prohibitum and thus morally neutral is flawed. As discussed above, mala prohibita and mala in se are not necessarily exhaustive categories. Professor Tony Honoré's writing helps to explain why these categories are not exhaustive. The view that tax evasion is morally neutral, except for its illegality, seems predicated on the assumption that we can always determine the moral status of an act independently of a community's laws.[120] According to this assumption we can determine whether an act is immoral through rational moral argument and need not draw any firm conclusions from what the formal legal system has to say on the matter.[121]

Like many unspoken assumptions, this one has prima facie plausibility. In a great many cases we do seem to know the moral value of an act independently of its legal status; we do not need to trawl through the statute books to know that murder, rape or theft are wrong. It is true that some conduct, for example tax evasion and avoidance, is impossible to evaluate, or even to define, if removed from its legal context. It begs the question, however, to assert that because such conduct relies on legal definition, its content is morally neutral and its only moral component is due to its legal status. Such an argument could only be logically sound if it first established that legally independent wrongfulness, in a mala in se sense, and moral content derived only from legal status, in a mala prohibita sense, exhaust the moral landscape.

For Honoré, a "picture of morality as a blueprint and law as a structure put up either according to or in disregard of it is ... misleading."[122] We live in collective groups or political communities. A large part of morality relates to questions of how to co-exist and to cooperate with others: "[t]he core of morality is, in a broad sense, political."[123] It follows that in a complex society, a viable morality necessarily has a legal component. That is, the morality of a complex society will be incomplete without a legal system containing certain types of laws.[124]

Broadly speaking, when we consider moral questions we are concerned with the ways in which our actions may significantly impact on other people, either

[120] Tony Honoré "The Dependence of Morality on Law" (1993) 13 Oxford J Legal Stud 1, 1.
[121] Honoré "The Dependence of Morality on Law" above, n 111, 1.
[122] Honoré "The Dependence of Morality on Law" above, n 111, 3.
[123] Honoré "The Dependence of Morality on Law" above, n 111, 2.
[124] Honoré "The Dependence of Morality on Law" above, n 111, 2.

individually or collectively. We are interested in how these impacts do, or should, limit or restrain our behaviour.[125] For Honoré, law relates to morality in two central ways. First, law can form a part of morality. Secondly, laws are open to moral criticism.[126] The second of these assumptions is fairly uncontroversial from a legal positivist standpoint; if there is no necessary connection between morality and law, then it is possible to criticise an existing law on moral grounds. However, the first assumption requires further explanation.

B Honoré: Some Aspects of Morality Depend on External Definition

Moral principles are normative in the sense that they tell us how we should act or behave. However, such principles are not always determinatively prescriptive. Moral principles are often quite general; while they provide a guide to the sorts of ways individuals should generally act, they do not always contain enough information to allow them to tell individuals exactly what is required of them in particular situations. For instance, there may be a general principle that individuals should be "generous" or "kind" or should "not always put their own interests first" but these principles are fairly open ended and vague. These broad duties are not always sufficient to define individuals' particular duties to each other when their interests conflict.[127] Particularly when a moral requirement is considered compulsory, it requires a measure of specificity so that individuals know just what they are morally required to do. If the core of morality is interacting with others, then our morality, as a set of general principles, is incomplete or lacks specificity. For morality to be complete and meaningful in practice it needs some additional definition from a source outside of itself.

C Primitive and Pre-Legal Societies

By Honoré's account, a "primitive" society is one without the formal legal institutions that are present in "complex societies."[128] In primitive, tribal, small-scale or pre-legal societies, individuals' moral duties to each other tend to be well defined. Individuals

[125] Honoré "The Dependence of Morality on Law" above, n 111, 2.
[126] Honoré "The Dependence of Morality on Law" above, n 111, 2.
[127] Robert Sugden *The Economics of Rights, Co-operation and Welfare* (Palgrave MacMillan, New York, 2004) 150.
[128] Honoré "The Dependence of Morality on Law" above, n 111, 2. Note, this distinction is similar to the one drawn by Popper between "closed" and "open" societies. See KR Popper *Open Society and Its Enemies: Volume 1 – The Spell of Plato* (2ed, Routledge and Kegan Paul, London, 1952) 172-90.

in a small-scale community tend to know each other or to be closely linked. [129] Their social roles are defined by fairly rigid customary or conventional practices and are clearly understood.[130] It is rare for an individual to doubt how to act in a certain situation as there will generally be clear social norms to which she understands she must adhere.[131] Despite the indeterminacy of general moral principles, it is unusual for individuals in a primitive society to experience moral conflict.

This determinacy of the social morality in a primitive society cannot stem solely from morality itself. The outside determinant is social convention.[132] Consider an interdependent community of one hundred individuals that does not have a legal system. If it functions as a cooperative society, its members must all know what community duties they owe. Human collective societies by nature appear to involve more than just the bare minimum level of cooperation necessary to, for example, provide shelter, avoid starvation and allow for reproduction.[133] In a primitive society, each individual's moral duties towards his fellows would be defined to a large extent by convention and enforced not necessarily through rules or laws but via some means of social pressure.[134]

A convention that might arise in such a society could be that each household contributes one able bodied man in wartime. Such a duty gives shape to the abstract duty to contribute to the protection of the community.[135] Conventions to take in orphaned relatives or to contribute food or shelter to sick members of the community similarly define the more nebulous duty to be generous to others.

D Modern Legal Societies

A complex, larger, modern society involves more independence and separation of individuals. While it is still a collective, the links between the members are less direct, more abstract.[136] The links between members of a complex society are too numerous and varied for convention to determine or settle the content of individuals' moral

[129] Popper, above, n 119, 176.
[130] Popper, above, n 119,172.
[131] Popper, above, n 119,172.
[132] Honoré "The Dependence of Morality on Law" above, n 111, 2.
[133] Geoffrey Sawer *Law and Society* (Oxford University Press, Oxford, 1965) 28; JC Smith *Legal Obligation* (The Athlone Press, London, 1976) 135.
[134] Honoré "Must We Obey?", above, n 110, 49; Sawer, above, n 124, 37.
[135] Honoré "The Dependence of Morality on Law" above, n 111, 7.
[136] Popper, above, n 119,174.

obligations to one another.[137] In a complex society there are more "others" to consider and a moral code needs to become even more specific.[138] Yet individuals tend not to feel that they have links to a greater number of "others", but as though, since there are so many seemingly faceless "others", they are not linked to any of them.[139]

Determination of vague moral norms is just as necessary in a complex society as in a primitive one, but convention and social pressure will not be sufficient to define nor to enforce particular moral requirements.[140] Honoré argues that law fulfils this determinative role in complex societies.[141]

E The Comparison of Primitive to Complex Societies.

Some might doubt whether the primitive pre-legal societies can be validly compared to modern legal societies and political communities;[142] perhaps there is a difference not only of scale and complexity but also of kind between the two social models. However, such an objection would have to be spelt out in more specific terms to be convincing. Generally, changes of scale need not affect the principles at work. It is hard to see why an individual's duties to her fellows in a small, close-knit group would cease merely because the group reaches such a size that she no longer has direct links with every member. Certainly, her duties take on a more abstract nature and may be less straightforwardly conceptualised by individuals, but they remain duties. Judicial statements that have reflected a belief in the moral dubiousness of tax avoidance often invoke such notions of duties of good citizenship and duties to other taxpayers.[143]

F Tax Evasion and the Middle Ground Between Mala in Se and Mala Prohibita: A Legally Constructed Moral Wrong?

If some moral obligations need outside determinants, then the problem with viewing mala in se and mala prohibita as exhaustive categories becomes clear. Legally defined crimes like tax evasion are not necessarily morally neutral.

In cases where either convention or law is required to give an abstract moral norm sufficient definition to be socially prescriptive, the defined moral obligation cannot be

[137] Tony Honoré "The Dependence of Morality on Law" above, n 111, 8.
[138] Petersen, above, n 77,16.
[139] Petersen, above, n 77, 3.
[140] Honoré "Must We Obey?", above, n 110, 49.
[141] Honoré "The Dependence of Morality on Law" above, n 111, 11.
[142] See Simmons *Moral Principles and Political Obligations*, above, n 109, 140.
[143] *Latilla v IRC*, above, n 39, 381 Lord Simon.

said to be universal or wholly or causally distinct from the convention of law that shapes it. Within the context of a modern, complex society such us our own, convention will be insufficient for this purpose, so we will necessarily be dealing with a moral duty defined by law. The abstract moral duty that relates to tax evasion is something like "to contribute to one's cooperative society". Taxation law gives shape to this moral duty by defining the measure of taxes on the forms of income that a taxpayer must pay.

Such a legally defined moral duty does not clearly fit within the malum in se category: tax is by definition something that is legally imposed; it is logically impossible to pay a tax if no there is no tax law to impose it; to fail to do something that is logically impossible cannot be immoral. However, while this thought experiment tells us that tax evasion is not malum in se, it does not establish that, given that we do have tax laws, there is not something wrong with evading them beyond the general obligation to obey the law. To class tax evasion as malum prohibitum and therefore morally neutral seems to fail to take proper account of the fact that, although legally defined, the duty is a moral one.

When tax evasion is understood this way, we can see first that it is morally wrong not only because it is illegal but also because, within our legal and societal context, our broad moral obligation to contribute to the collective has taken the specific shape of a duty to pay our taxes. Tax evasion is thus a wrong in a deep sense. Given that tax *evasion* is morally wrong in virtue of its content as well as its legal status, it is not convincing to suggest that the mere fact that tax is not illegal means that it is also not immoral. From this perspective, the factual similarity between avoidance and evasion and the fineness of the legal line between them suggest the opposite conclusion from the one drawn by the judges. If tax avoidance is factually almost indistinguishable from tax evasion, and if despite being a legal construct tax evasion is in a deep sense immoral, act, then tax avoidance is similarly immoral.

VII CONCLUSION

Judges and commentators have made much of the legal difference between evasion and avoidance. Many judges have considered that while evasion is immoral because it is in breach of the law, legal avoidance involves no such moral misstep. However, it is a logical confusion to draw moral conclusions from solely legal observations. The fact that conduct is not illegal does not necessarily mean that it is also moral. For

instance, it is legal to sell cigarettes, but this observation does not tell us whether selling cigarettes is moral. It is true that legal conduct cannot be immoral in one special sense: that is, since it is not illegal, it cannot be immoral for the reason of illegality. However, conduct can be immoral for many other reasons than that it breaches the law. Evasion is immoral for more than just legal reasons.

The difference between evasion and avoidance is a matter of law, not of relevant fact. The difference cannot tell us anything about the differences or similarities between evasion and avoidance in moral terms, except for the special case that evasion is immoral anyway to the extent it is illegal. Judges overlook the considerable factual similarity between avoidance and evasion: both aim to minimise tax liability and both have similar economic effects. Judges are correct when they note that there is a legal dissimilarity between the evasion and avoidance. However, the factual similarities are of much greater moral significance than the legal difference; from the perspective of morality we should look at the two as phenomena as being just one phenomenon.

LAW IN MODERN SOCIETY[1]

Roberto M. Unger

An understanding of liberal society illuminates, and is illuminated by, an awareness of that society's legal order and legal ideals. For the rule of law has been truly said to be the soul of the modern state. The study of the legal system takes us straight to the central problems faced by the society itself.

If this hypothesis, which underlies my argument, is correct, then any revision of the nature and uses of law will reveal changes in the basic arrangements of society and in men's conceptions of themselves. At the same time, whatever we can learn about these social changes will help us reinterpret the transformation of the legal order. In this spirit, I discuss on the following pages some aspects of the way certain countries, the Western capitalist social democracies, have become postliberal societies.

The characteristics of these societies undermine the rule of law and they strengthen tendencies in belief and organization that ultimately discourage reli-

ance on public and positive rules as bases of social order. These startling trends will force us to reexamine our view of the situation and the prospects of postliberal societies. For my immediate purposes, it is enough to emphasize two commonly observed sets of features of this novel form of social life.

The first group of features refers to the overt intervention of government in areas previously regarded as beyond the proper reach of state action. The response to the problems of unjustified hierarchy, a response the rule of law failed to provide, is now sought from the government. The rank order itself increasingly moves to the centre of political debate and political action. As the state becomes involved in the tasks of overt redistribution, regulation, and planning, it changes into a welfare state.

The other notable set of attributes of postliberal society is but the reverse side of the events just enumerated : the gradual approximation of state and society, of the public and the private sphere. For one thing, the state's pretense to being a neutral guardian of the social order is abandoned. For another thing, private organizations are increasingly recognized and treated as entities with the kind of power that traditional doctrine viewed as the prerogative of government. People may become more conscious of what was always partly true, though perhaps less so in earlier periods : society consists of a constellation of governments, rather than an association of individuals held together by a single government. The state that has lost both the reality and the consciousness of its separation from society is a corporate state.

Now let us see how these welfare and corporatist tendencies affect the society's normative order.

Welfare state developments influence the legal order of postliberal society in a variety of ways. But two kinds of immediate influence seem particularly significant.

The first type of effect is the rapid expansion of the use of open-ended standards and general clauses in legislation, administration, and adjudication. For example, the courts may be charged to police unconscionable contracts, to void unjust enrichment, to control economic concentration so as to maintain competitive markets, or to determine whether a government agency has acted in the public interest. Such indeterminate prescriptions have always existed in the law, but they grow rapidly in prominence because of the transformations to which I refer.

The second major impact of the welfare state on law is the turn from formalistic to purposive or policy-oriented styles of legal reasoning and from concerns with formal justice to an interest in procedural and substantive justice. Before further discussion, these terms should be defined.

Legal reasoning is formalistic when the mere invocation of rules and the deduction of conclusions from them is believed sufficient for every authoritative legal choice. It is purposive when the decision about how to apply a rule depends on a judgment of how most effectively to achieve the purposes ascribed to the rule. The difference between these two types of legal reasoning is one between the criteria thought appropriate to the overt justification or criticism of official decisions ; it does not pretend to describe the actual causes and motives of decision.

An ideal of justice is formal when it makes the uniform application of general rules the keystone of justice or when it establishes principles whose validity

is supposedly independent of choices among conflicting values. It is procedural when it imposes conditions on the legitimacy of the processes by which social advantages are exchanged or distributed. It is substantive when it governs the actual outcome of distributive decisions or of bargains. Thus, in contract law, the doctrine that bargains are enforceable given certain externally visible manifestations of intent exemplifies formal justice ; the demand that there be equality of bargaining power among contracting parties illustrates procedural justice ; and the prohibition of exchanges of two performances of unequal value, however value may be assessed, represents substantive justice.

A formal view of justice requires, to be coherent, a belief in the possibility of formalistic legal reasoning. And it is likely to be most persuasive in the realm of exchanges among individuals rather than in that of governmental distribution, which inevitably involves choices among conflicting interests. Thus, it tends to distinguish sharply between an impersonal justice of reciprocity that dispenses with distributive premises and an arbitrary justice of distribution whose pronouncements are never impartial and general enough to have anything more than the appearance of law.

Procedural or substantive notions of justice become important as purposive forms of legal reasoning are adopted, and they in turn give impetus to those varieties of argument. For policy-oriented legal discourse forces one to make explicit choices among values, and the pursuit of procedural or substantive justice requires that rules be interpreted in terms of ideals that define the conception of justice. Hence, every decision about the principles that govern exchange is seen to rest upon procedural or distributive premises and to have procedural or distributive consequences.

Postliberal society witnesses an escalating use of open-ended standards and a swing toward purposive legal reasoning and procedural or substantive approaches to justice. This is a change of emphasis rather than a sequence of clearly differentiated stages. In few societies have these shifts followed a line of uninterrupted progression. Periods of greater stress on formalistic legal reasoning and formal justice have followed eras of a more policy-oriented mode of legal discourse, as in nineteenth-century America. Even during the hegemony of formalism, there has often been a widespread awareness of the fact that the legal order was redistributing resources among groups and classes. Nevertheless, I shall argue later that present-day tendencies differ from their earlier counterparts not merely because of the more pronounced, persistent, and universal character of the contemporary developments, but, above all, because of the emergence in the welfare-corporate state of a unique relationship among problems of formality, equity, community, and equality in law.

The immediate causes of the postliberal moves toward purposive legal reasoning and procedural or substantive justice are directly connected with the inner dynamic of the welfare state. These moves appear as ways to deal with concentrated power in the private order or to correct the effects of a system of formal rules. As government assumes managerial responsibilities, it must work in areas in which the complexity and variability of relevant factors of decision seem too great to allow for general rules, whence the recourse to vague standards. These standards need to be made concrete and individualized by persons charged with their administrative or judicial execution.

The reasons for the greater emphasis on purposive legal reasoning and on procedural or substantive justice are more obscure and less amenable to a comprehensive interpretation. Changes in the theoretical understanding of language, in the character of common beliefs about the basis and scope of legitimate state action, and in the structure of the rank order all seem to play a part. Language is no longer credited with the fixity of categories and the transparent representation of the world that would make formalism plausible in legal reasoning or in ideas about justice. In the absence of belief in the naturalness of existing hierarchies of power or distribution, the legitimacy of governmental, including judicial, activity comes to depend increasingly on the welfare consequences of that activity. Finally, the vicissitudes of class struggle strip the state of every pretense to impartiality and transform it into an acknowledged tool of factional interest in a social situation in which the dictates of justice are still believed to be unknowable.

Whatever the causes of the trends I have described and however much they may vary from one country to another, their chief effects on the law seem clear. They repeatedly undermine the relative generality and the autonomy that distinguish the legal order from other kinds of law, and in the course of so doing they help discredit the political ideals represented by the rule of law.

Open-ended clauses and general standards force courts and administrative agencies to engage in ad hoc balancings of interest that resist reduction to general rules. One of the corollaries of generality in law is a severe limitation of the range of facts considered relevant to the making of official choices. If the number of pertinent factors of decision is too large, and each of them is constantly shifting, then categories of classification or criteria of analogy will be hard to draw and even harder to maintain. But the kinds of problems to which comprehensive standards characteristically apply tend to defy such limitations. They involve the conflict of numerous and inchoate interests against the background of a refusal to sacrifice any one of these interests completely to the others.

When attempts are made to codify standards, to reduce them to rules, their character is distorted. Either a large area of uncontrolled discretion and individualization subsists under the trappings of general norms, or the flexibility needed to make managerial decisions or to produce equitable results is lost. The same dialectic of illusion and petrification can be observed in the analagous processes by which Roman praetorian law was overtaken by imperial legislation, English equity lost out to common law, and the customary or sacred laws of non-Western societies were codified by colonial administrators.

Purposive legal reasoning and nonformal justice also cause trouble for the ideal of generality. The policy-oriented lawyer insists that part of interpreting a rule is to choose the most efficient means to the attainment of the ends one assigns to it. But as the circumstances to which decisions are addressed change and as the decisionmaker's understanding of the means available to him varies, so must the way he interprets rules. This instability of result will also increase with the fluctuations of accepted policy and with the variability of the particular problems to be resolved. Hence, the very notion of stable areas of individual entitlement and obligation, a notion inseparable from the rule of law ideal, will be eroded.

The quest for substantive justice corrupts legal generality to an even greater degree. When the range of impermissible inequalities among social situations ex-

pands, the need for individualized treatment grows correspondingly. No matter how substantive justice is defined, it can be achieved only by treating different situations differently. Thus, for example, it may become necessary to compensate for an existing inequality with a reverse preference afforded by the legal order to the disadvantaged group. Priorities among groups in turn shade imperceptibly into preferences among individuals and individual situations.

The history of the law of obligations and of liability rules in many Western social democracies illustrates another way in which the insistence on substantive justice enters into conflict with established notions of generality. Classic theories of contractual and delictual liability drew a sharp line between the allegedly impersonal justice of reciprocity; with which they were concerned, and distributive justice, which, if it existed at all, was the province of politics and the marketplace. At the same time, they confined liability to areas of conduct that seemed amenable to general rules and they asserted its absolute character within those confines.

In the era of the welfare state and of policy-oriented legal discourse, there is a firmer recognition that exchange rules do have a distributive significance. Nonetheless, the attempt to take distributive criteria into account in an adjudicative setting unavoidably forces the courts into fields in which the complexity of relevant factors and the lack of widely shared standards of justice make generalization hard to come by and to stick with. The situation is aggravated by the impulse to extend liability, in response to equitable considerations, to areas where the same sorts of problems arise. And the difficulty is compounded still further by the willingness, in criminal as well as in private law, to admit a growing list of exculpatory conditions within this enlarged sphere of liability. For the granting of an excuse turns on judgments about particular persons and individual situations, judgments that resist statement as rules.

The same events that subvert the generality also tend to destroy the relative autonomy of the legal order in its substantive, methodological, institutional, and occupational dimensions.

Overarching standards invite their appliers to make use of the technician's conception of efficiency or the layman's view of justice. If, for example, one seeks to give content to the conception of good faith in contract law, one must go outside the narrow confines of lawyers' learning to consult the practices and enter into the thought patterns of a certain social group.

As purposive legal reasoning and concerns with substantive justice begin to prevail, the style of legal discourse approaches that of commonplace political or economic argument. All are characterized by the predominance of instrumental rationality over other modes of thought. Indeed, policy-oriented legal argument represents an unstable accommodation between the assertion and the abandonment of the autonomy of legal reasoning, just as a procedural justice mediates between formal and substantive justice.

The decline in the distinctiveness of legal reasoning is connected with the need administrators and judges have of reaching out to the substantive ideals of different groups, of drawing upon a conventional morality or a dominant tradition. These changes in the substance and method of law also help undercut the identity of legal institutions and of the legal profession. Courts begin to resemble openly first administrative, then other political institutions. Thus, the difference between lawyers and other bureaucrats or technicians starts to disappear.

The cumulative impact of the movements discussed in the preceding pages is to encourage the dissolution of the rule of law, at least insofar as that form of legality is defined by its commitment to the generality and autonomy of law. To be sure, autonomy and generality could never be meant as completely actualized descriptions of the legal order in liberal society; they are no more than ideals which the liberal form of social life makes necessary to entertain and impossible to achieve fully. What distinguishes the law of the postliberal period is primarily the turning away from these ideals, a change of course that, despite its apparent insignificance, indicates important shifts in human belief and social order.

The corporatist tendencies of postliberal society have potentially an even more dramatic effect on the law than the welfare state trends. If the latter contribute to the disintegration of the rule of law, the former ultimately challenge the more universal and elementary phenomenon of bureaucratic law, law that is public and positive.

The spearhead of corporatism is the effacement both in organization and in consciousness of the boundary between state and society, and therefore between the public and the private realm. As the state reaches into society, society itself generates institutions that rival the state in their power and take on many attributes formerly associated with public bodies. It is doubtless true that much of the earlier separation of government and society may have been more a matter of vision than of reality. But here one must tread carefully. The images people hold of their social situation are an integral part of those situations; indeed, they establish their specifically social meaning. Thus, a modest change of emphasis in forms of organization may be important if it is accompanied by a transformation of belief. The corporatist developments, like the welfare state ones mentioned earlier, seem to exemplify this principle.

Corporatism's most obvious influence on the law is its contribution to the growth of a body of rules that break down the traditional distinction between public and private law. Thus, administrative, corporate, and labour law merge into a body of social law that is more applicable to the structure of private-public organizations than to official conduct or private transactions. But though this development undermines the conventional contrast of public and private law, it does not necessarily destroy the broader difference between the law of the state and the internal, privately determined regulations of private associations. Insofar as private law is laid down by the state, it too is, in this more comprehensive sense, public.

The deepest and least understood impact of corporatism is the one it has on the very distinction between the law of the state and the spontaneously produced normative order of nonstate institutions. As private organizations become bureaucraticized in response to the same search for impersonal power that attracts government to the rule of law principle, they begin to acquire the features, and to suffer the problems, of the state. At the same time, the increasing recognition of the power these organizations exercise, in a quasi-public manner, over the lives of their members makes it even harder to maintain the distinction between state action and private conduct. Finally, the social law of institutions is a law compounded of state-authored rules and of privately sponsored regulations or practices; its two elements are less and less capable of being separated. All these movements, which tend to destroy the public character of law, carry forward a

process that begins in the failure of liberal society to keep its promise of concentrating all significant power in government.

The tendency of large corporate organizations to become bureaucratized and to produce a body of rules with many of the characteristics of state law should not be confused with an increasing regulation of the corporation by the state. In fact, quite the opposite may be true : the bureaucratization of corporate institutions may be associated with their ability to become relatively independent power centres with decisive influence over government agencies.

Corporatist tendencies are often associated with demands for the change of public and private organizations into democratic communities. These demands are still usually encountered as ideologies rather than as institutional realities. Nonetheless, they are just as much rooted in the structure of the postliberal order as the rule of law ideal is in the nature of liberal society.

Sometimes the communitarian aspirations are part of a radical attack on the corporate bodies. At other times, they present themselves under the guise of a reformist politics of participation. But whatever their immediate source or objective, they all betray a dissatisfaction with the nature of hierarchy and therefore of personal existence under liberalism. They all attempt to show how the fundamental experience of unjustified power and arbitrary consensus can be dealt with when the rule of law fails to dispose of it. And they all look for an alternative to the ideal of legality in the notion of a community bound together by a shared experience and capable of developing its own self-revising customs or principles of interaction. The profound and irreconcilable differences between rightist and leftist interpretations of the communitarian program have to do with the extent to which they envisage the community as arising from a preservation and strengthening, or from a destruction, of the rank system of liberal society.

An integral part of the search for community in both its conservative and its revolutionary variant is the aim of avoiding the manipulation of social life by imposed rules and of respecting the spontaneously produced internal customs of each communal group. This longing is heralded by the currents in social thought and jurisprudence that emphasize the "living" or "inner" law of associations in contrast to the made rules of the state. What is ultimately at issue is therefore the positive character of law itself : whether or not significant reliance will be placed upon made and articulated rules as opposed to immanent and implicit custom. And behind this conflict of types of law lies a more general antagonism between forms of social life — one for which order is a spontaneous byproduct of interaction ; another for which it represents authority imposed from above or outside. [pp. 192-203]

NOTE

1. R.M. Unger, *Law in Modern Society* (New York : The Free Press, 1976), pp. 192-203.

Should Tax Legislation be Written from a Principles and Purpose Point of View or a Precise and Detailed Point of View?

JOHN PREBBLE*

Introduction

PEOPLE have been concerned about the increasing detail of tax legislation for much of the history of income tax law. In 1936, for example, the United Kingdom Codification Committee said:

> "[T]he statutes of 1842 and 1853 were relatively simple. The growth of legislation since 1907 and its increasing complexity have been in large measure due to the high rates of tax in operation. ... The space occupied by the provisions relating to ... reliefs and exemptions is now prodigious, and contrasts with the comparative brevity of the earlier code...".[1]

No doubt the high rates of tax had something to do with the developments deplored by the Codification Committee. But a more important reason is surely the steadily increasing sophistication of income tax legislation. Early income tax law in most countries was crude, and riddled with loopholes.[2] Tax policy makers have come to understand more thoroughly the principles of economics as they apply to income taxation (the time value of money being a prominent example) and the ways in which business and investment can be structured to avoid income tax laws that are rudimentary in form. Policy makers have responded by urging on parliaments rules that, while increasingly complex, have had the merit of immensely strengthening the income tax base and have promoted horizontal equity by closing off planning opportunities available to some taxpayers but not to others.

Thesis

Income tax law is different in kind from law in general, and from tax laws that depend on transactions or on states of fact. Unlike other laws, income tax law does not relate directly to its subject matter, which is the facts and legal relationships of business activity. Income tax law ignores some facts and transactions and it recharacterises others. In other words, tax law is dislocated from its subject matter.

* BA, LL.B (Hons) (Auckland); BCL (Oxon.); JSD (Cornell); Inner Temple. Barrister and Professor of Law, Victoria University of Wellington, New Zealand. This article has been revised from papers prepared for the New Zealand Inland Revenue Department's Tax Drafting Conference, Auckland November 27–29, 1996 and for addresses to the Australian Branch of the International Fiscal Association in Sydney and Melbourne on July 28 and 30, 1997.

1 *Report of the Income Tax Codification Committee*, Cmd. 5131, para. 18, quoted by J. Avery Jones in "Tax Law: Rules or Principles".

2 See, *e.g.* J. Prebble (1996) 17 *Fiscal Studies* 63; [1996] B.T.R. 580 "100 Years of Income Tax" (1993) 47 *Bulletin for International Fiscal Documentation* 59.

[1998] B.T.R.: No.2 © SWEET & MAXWELL AND CONTRIBUTORS 1998

This separation of income tax law from its subject matter is a crucial and ever-present factor to be borne in mind when one tries to answer the question that is the subject of this article, "Should tax legislation be written from a principles and purpose point of view or a precise and detailed point of view?" The thesis of this article is that the answer to the question is that "precise and detailed" must win out in the end, though "principles and purpose" drafting has an important role.

Meaning of "principles and purpose"

For present purposes, a very useful rephrasing of the question, principles and purpose or precise and detailed, was given recently by John Avery Jones, then joint editor of the *British Tax Review*, in the first sentence in this passage from his 1996 Institute for Fiscal Studies Annual Lecture:

> "Could our tax legislation be rewritten so that it would be construed in accordance with principles rather than containing nothing but rules trying unsuccessfully to cover every eventuality?[3] Ronald Dworkin makes the distinction, which I think we are searching for here, between rules and principles.[4] The distinction is that rules are applicable in an all-or-nothing fashion, whereas principles are not. Principles can have exceptions, can conflict with one another; can, for example, apply only when conduct is reasonable; and they can give one guidance about how to deal with the points not expressly covered by the law. Rules do not conflict with each other (if they do, one of them must give and become a subsidiary rule). Principles do not conflict with rules. If a rule is clear, the rule applies and that is the end of it, even though it is an exception to the principle and therefore in conflict with it. But the real use for the principle is to determine what the rule means in the first place (and hopefully to reduce the amount of detail required in stating the rule), in which case there is no conflict."[5]

As indicated, the thesis of this article is that generally speaking tax legislation could not successfully be rewritten in accordance with principles only, rather than having rules. Mr Avery Jones is correct, however, in identifying the usefulness of principles as a key to interpreting rules.

Ectopia

Income tax law poses problems that are peculiar to it, and that distinguish income tax law not only from other law in general, but also from law that applies to other types of taxation. The author has elsewhere described these problems as "ectopia", meaning "dislocation", and has described the features and significance of ectopia in income tax law in a number of articles.[6] The utility of the concept as an analytical tool has received some support. Lord

[3] Here, Avery Jones asks: "Or are there no principles in tax law?", as suggested by John Prebble in "Why is Tax Law Incomprehensible?" [1994] B.T.R. 380.

[4] *Taking Rights Seriously* (Duckworth, 1977) p. 22.

[5] J. Avery Jones "Tax Law: Rules or Principles" (1996) 17 *Fiscal Studies* 63, 75–76.

[6] "Ectopia, formalism, and anti-avoidance rules in income tax law" (1994) in W. Krawietz N. MacCormick & G. H. von Wright (eds) *Prescriptive Formality and Normative Rationality in Modern Legal Systems, Festschrift for Robert S. Summers* (Duncker and Humblot, Berlin), at pp. 367–383; "Philosophical and design problems that arise from the ectopic nature of income tax law and their impact on the taxation of international trade and investment" (1995) 13 *Chinese Yearbook of International Law and Affairs*, 111–139, reprinted [1997] B.T.R. 383; "Can income tax law be

[1998] B.T.R.: No.2 © SWEET & MAXWELL AND CONTRIBUTORS 1998

Cooke mentioned it in his first Hamlyn Lecture on November 7, 1996,[7] and in his dissenting judgment in *Wattie v. Commissioner of Inland Revenue*[8] in the New Zealand Court of Appeal Thomas J. thought that the concept shed some light on the difficulty of distinguishing income from capital. A short introduction to what is meant by the ectopia of income tax law may suffice here. Ectopia is at its most acute in respect of business income, which is the primary subject of this article.

The chief problem with income tax law is that it depends on the concept of income, which is an artificial construct. "Income", meaning the profits of a business for tax purposes, is not something that one can see in nature. Contrast, for example, a transaction, or an invoice, which are essential keystones of a value added tax. "Income" as used in tax legislation is not even an abstract concept that has an independent existence, such as anger, or *mens rea*, or anti-competitive activity.

"Income", meaning "profits of a business", does of course exist in nature. It is possible to calculate the profits of a business by referring to identifiable facts, the most important being the receipts and expenditure of the business over its lifetime. But the result of this calculation is of secondary importance for income tax law. Income tax law requires us to allocate profits to periods. (A year is invariably chosen, though there is no inexorable reason why that should be so.) That income is a construct or conception rather than a fact that has an independent existence was recognised by Dixon J. in *Carden's Case*, where he said:

> "Income, profits and gains are *conceptions* of the world of affairs and particularly of business ... in nearly every department of enterprise and employment the course of affairs and practice of business have developed methods of estimating and computing in terms of money the result over an interval of time produced by the operations of business, by the work of the individual, or by the use of capital. The practice of these methods of computation and the general recognition of the principles upon which they proceed are responsible in great measure for the *conceptions* of income, profit, and gain and, therefore, may be said to enter into the determination or definition of the subject which the legislature has undertaken to tax." (Emphasis added.)[9]

Dixon J. was referring to business people's concept of income, but his words are equally apt to describe the concept of income of tax law, which is essentially a refinement of the concept of income of a business. This refinement sometimes leads to "a degree of tension between ... formal legal doctrine and ... a computation of taxable profits which is wholly divorced from commercial reality."[10]

Time

The need to divide income into sections by reference to the "intervals of time" mentioned by Sir Owen Dixon causes us to treat many transactions in a counterfactual or counterlegal

simplified?" (1996) 2 *N.Z. Journal of Taxation Law and Policy* 187; see also "Why is tax law incomprehensible?" [1994] B.T.R. 380–393.

[7] Rt. Hon. Lord Cooke of Thorndon, "A real thing" in *Turning Points of the Common Law* (1997) p. 12.

[8] (1997) 18 N.Z.T.C. 13,297, 13,309–13,310.

[9] *C of T (South Australia) v. Executor, Trustee, and Agency Co. of South Australia Ltd* (1938) 63 C.L.R. 108, 152, quoted in *Commissioner of Inland Revenue v. Mitsubishi Motors New Zealand Ltd* [1995] 3 N.Z.L.R. 513 at 515, PC.

[10] *ibid.* 517.

[1998] B.T.R.: No.2 © SWEET & MAXWELL AND CONTRIBUTORS 1998

manner. For example, for most businesses the cost of stationery is an expense that is deductible in calculating taxable income. But suppose that in the current year I buy a large supply of envelopes, and I have many envelopes left over at the end of the year. My business uses these leftover envelopes next year and the year after. Section EF1 of the New Zealand Income Tax Act 1994 forbids me to deduct the cost of the leftover envelopes in the current year, when I bought them. I must instead apportion the cost partly to the accounts of the next two years. From an economic and tax policy point of view this rule is entirely appropriate. But from a factual and procedural point of view, section EF1 apportionment requires a taxpayer to treat certain expenditure as if it were income.

A counterlegal example is seen in section FC10 of the Act, which deals with hire purchase agreements. Essentially, section FC10 treats hire purchase agreements as a sale of the item that is subject to the agreement, accompanied by a loan of the purchase price. Sections FC5, 6, and 7 have a similar effect in respect of certain leases. Again, these provisions are entirely consistent with policy; but that policy requires income tax law to treat hire purchase agreements and some leases as if they were legally transactions of a different kind. These examples may demonstrate that there is often a dislocation between income tax law and either or both of the factual and legal objects that are the subject of that law.

Place

Income tax law also requires us to identify income as coming from a particular source, and to identify that source with a particular physical fact, namely, a taxing jurisdiction. Yet this exercise is a logical impossibility.

We can say that a river, which is a physical fact, has its source in a particular country, which is another physical fact. But income is the difference between a taxpayer's receipts and expenses, subject to rules that locate these receipts and expenses in time. Logically, income cannot have a source that is located according to the principles of geography. Yet we must treat income as if it does have a physical source. Indeed, in the leading case of *Nathan v. Federal Commissioner of Taxation* Isaacs J. described the source of income as "a practical, hard matter of fact".[11] His Honour would have been more accurate to say that income tax law *requires us to treat* the question of source as a hard, practical matter of fact rather than that the source of income is such a fact.

Companies

A company is a convenient fiction,[12] invented to enable people to band together for business or investment, and to enable management to be separated from capital. The corporate form works very well when it is measured against these objectives. As between capitalists and creditors, for example, the fiction that a company exists as a real entity separate from its members and employees works reasonably well.[13] But when we add a third party, the state in its taxing role, all sorts of problems emerge.

[11] (1918) 25 C.L.R. 183 at 189–190.

[12] J. W. Salmond, *Jurisprudence* (1902) p. 335, *ibid.* (7th ed.) (1924) p. 330. See further J. Prebble "Can income tax law be simplified?" (1996) 2 *N.Z. Journal of Taxation Law and Policy* 187, para. 2.6. For a summary of competing theories see A. Frame *Salmond, Southern Jurist* (Wellington 1995) pp. 88–71.

[13] *Salomon v. Salomon & Co.* [1897] A.C. 22, *pace* O. Kahn-Freund, "Some reflections on company law reform" (1944) 7 M.L.R. 54.

[1998] B.T.R.: No.2 © SWEET & MAXWELL AND CONTRIBUTORS 1998

Companies and their members or associates; as legally independent persons, can organise their affairs so as to frustrate the state when the state comes along to levy income tax in respect of transactions that have already occurred. On the other hand, the corporate form can cause a company and its members to pay tax twice or more on the same income. To remedy these problems income tax legislation is apt to contain rules that expressly or by implication look through corporations to their shareholders. The first kind of problem provokes parliaments to enact anti-avoidance provisions such as controlled foreign company[14] and deemed dividend rules.[15] The second kind of problem may be remedied by, *inter alia*, imputation regimes[16] or by taxing companies as partnerships.[17]

All of these sorts of regimes involve treating companies as if they were fictions, but as if they were a different kind of fiction from the fiction that the rest of the legal system recognises. That is, there is in this respect a dislocation, or ectopia, between the rules of income tax law and the rules of the rest of the legal system. In another sense, these kinds of rules reduce ectopia in that they bring the economic and substantive focus of income taxation closer to its ultimate target, the individual. But this is achieved at the expense of setting aside legal relationships.[18]

Residence

When international questions are relevant there is a further layer of ectopia. In income tax systems that tax on the basis of residence or citizenship (and nearly all systems, or perhaps all of them, do so) it is necessary to ascribe a notional residence or citizenship to companies. There is an unreality about this process that is superficially obvious enough, but that becomes even more divorced from facts and fictions that exist in the world or in the legal system when one bears in mind that companies are themselves already artificial, fictitious creations.

Despite these considerations, a legal system that employs residence or citizenship as a connecting factor that triggers liability to income tax must treat a company as if it had a residence in as real a sense as an individual may be said to have a residence.

Difference from other laws

These elements of time and place (the latter in respect of both source and residence), exacerbated when there are companies involved, are the primary factors that cause income tax law to be ectopic. In contrast, the rest of law is in general indifferent in this respect. Most laws try to reflect as closely as possible the legal or physical facts, or fictions, to which they relate. Criminal law tries to describe actions that are proscribed as accurately as drafters can manage. Contract law does its best to describe an agreement that amounts to a legally enforceable bargain. Testamentary law says what it means by the term "will".

Most significantly in the present context, value added tax and death duty law describe the transactions or states of fact that cause these taxes to bite (in the latter case, death and

[14] *e.g.* New Zealand Income Tax Act 1994, subparts CG and MF.
[15] *e.g ibid.* s.CF2.
[16] *e.g. ibid.* subpart ME.
[17] *e.g. ibid.* subpart HG.
[18] See J. Prebble "Can income tax law be simplified?" (1996) 2 *N.Z. Journal of Taxation Law and Policy* 187, paras 3.3, 3.4.

the existence of wealth). In contrast, income tax law must often treat its subject matter as if it were factually or legally different from what it is.

Principles as an aid to interpretation

There can be no doubt that a statement of principle at the beginning of a series of rules will often be very helpful in interpreting the rules. A good example is section FC9 of the New Zealand Income Tax Act 1994, which reads:

> "This section and section FC10 are intended to result in hire purchase agreements . . . that are made in relation to personal property other than livestock or bloodstock being treated for the purposes for this Act in similar manner to a sale of the property with a loan of the purchase price being made by the person providing the finance under the agreement to the person obtaining under the agreement the user right to use the property."

Section FC10 is calculated to overcome the problems that hire purchase agreements can pose for income tax law. People are apt to focus on the passage of property under a hire purchase agreement, and to say that a vendor's profit does not arise until property has passed to the buyer. Since hire purchase agreements provide that property does not pass until the end of the term of the agreement, this approach defers the recognition of profit for some time. If people adopt this approach one result is that for tax purposes hire purchase agreements are treated quite differently from credit sales, under which profit is recognised at the time of sale.[19] Such a difference cannot be justified, considering that the two kinds of transaction are economically equivalent, and differ only in respect of their security mechanisms.

While section FC10 is not remarkably complex, consideration of some of its provisions shows the merit of explaining in section FC9 just what it is that section FC10 tries to achieve. Take section FC10(2) by way of example:

> ". . . if on or after the termination or expiry of a hire purchase agreement the lessee . . . does not acquire ownership of the hire purchase asset the subject of the hire purchase agreement,—
> (a) The lessor shall be deemed to have acquired ownership of the hire purchase asset from the lessee for an amount equal to the lessor's outstanding balance in relation to the hire purchase agreement; . . . on the date of the termination or expiry of the hire purchase agreement."

Having just read section FC9, the reader is better equipped to come to grips with the terms and meaning of section FC10.

It is noteworthy that, although section FC9 is expressed in terms of a principle, the principle is fairly narrow; it applies only to hire purchase agreements. Section FC9 could have achieved the same result with the generalisation, "All transactions are assessable according to principles of economic equivalence." But while this much broader principle would deal appropriately with hire purchase transactions, it is far too wide for general application throughout the income tax system.

[19] *Gardner, Mountain & D'Abrumenil Ltd v. I.R.C.* [1947] 1 All E.R. 650, HL; *J. Rowe & Son Pty Ltd v. FCT* (1971) 124 C.L.R. 421; *C.I.R. v. Farmers' Trading Co. Ltd* [1982] 1 N.Z.L.R. 449, CA.

Existing principles in income tax legislation

A significant part of many income tax regimes is framed in terms of rules that, on examination, turn out to be principles in disguise. Probably the most important of these rules are rules that draw a line between income and capital. For example, section BD2(2)(e) of the New Zealand Income Tax Act 1994, as amended by section 6 of the Taxation (Core Provisions) Act 1996, forbids a deduction in calculating taxable income for expenditure "of a capital nature", except where expressly allowed.

On the receipts side, there is no specific reference to capital. Section BD1(1) simply says that:

> "An amount is gross income of a taxpayer if it is included in the taxpayer's gross income under Part C (Income Further Defined), D (Deductions further defined), E (Timing of Income and Deductions), F (Apportionment and Recharacterised Transactions), G (Avoidance and Non-Market Transactions), H (Treatment of Net Income of Certain Entities) or I (Treatment of Net Losses)."

Among the more important items from the various lists that section BD1(1) refers to are sections CD3 and CD5, both inserted by section 26 of the 1996 Act:

> "CD3 **Business**—The gross income of any person includes any amount derived from any business.
> CD5 **Gross income according to ordinary concepts**—The gross income of a person includes any amount that is included in gross income under ordinary concepts." [*sic.* "Pursuant to ordinary concepts of income"?]

Although none of sections BD1, CD3, nor CD5 mentions "capital" taxpayers can be confident that Parliament did not intend to tax capital gains except where the Act expressly says so, or where the Act has that effect by necessary implication. Throughout the history of income taxation, a great deal of the time of the courts, of revenue officials, and of tax advisers has been occupied with drawing the line between capital and revenue, in respect of both receipts and expenditure. This huge and continuing exercise furnishes a sobering example of the way the law might have to work if, elsewhere in the Act, rules were replaced with principles. Consideration of that hypothetical exercise leads to two questions. First, is this a good model for other areas of the law? Secondly, is there anything peculiar to the capital/revenue distinction that mandates an approach based on a statement of principle, rather than on rules?

The peculiarity of the capital/revenue distinction

To answer the latter question first, yes, there is something peculiar about the capital/revenue distinction, in both senses of the term "peculiar". The capital/revenue distinction is one of the chief manifestations of ectopia in income tax law. The distinction is not based on any *a priori* defensible economic principle, nor on any other principle that has an existence independent of tax law. The distinction arises because, for administrative purposes, receipts and expenditure must be grouped into annual clumps, so that the net result can be calculated and taxed. Items that relate to a number of years cannot be conveniently grouped with revenue items, so they are labelled "capital" and either exempted from the tax system (in New Zealand) or dealt with under other rules (in many other countries with developed tax systems).

[1998] B.T.R.: No.2 © SWEET & MAXWELL AND CONTRIBUTORS 1998

The result is that it is hard to draft a principle, or even a principled rule, that can distinguish between capital and revenue. In fact, if we take the view that a normative statement does not deserve to be called a "principle" unless it is logical and coherent and can be reconciled with other norms that are generally accepted to be "principles", it is perhaps a mistake to suggest that sections BD2(2)(e), and BD1, CD3, and CD5, are expressions of principle at all.

They are not expressions of true principle, but they masquerade as such, in that they do not contain detailed rules to cover the very large area of income tax law that they govern. It has been left to the courts to develop this law over the last 100 years or so, as if it were a field of common law.

To turn to the first of the two questions asked earlier, is all this judicial activity a good model for other areas of the tax law? It would be rash to answer yes. Even if one thought that the line that separates capital and revenue may be found tolerably well these days, that is the result of many decades of case law, and, at that, case law that established its fundamental principles years before income taxation became as important as it is nowadays. In fact, even now the line between capital and revenue remains blurred, and, to take one important example, the New Zealand investment fund and superannuation industries have been in a state of uncertainty for some years as the courts seem to increase the area of revenue at the expense of the area of capital.[20] People planning for their retirement have to conclude that if they invest personally in the sharemarket increases in value of shares will be capital on sale, but if they delegate the task to a superannuation or investment fund manager such increases will be taxed as income.

This is not necessarily to say that the capital/revenue distinction should be marked out by bright line rules rather than left as it is. It may be that the unprincipled nature of the distinction means that the present solution is about the best that can be hoped for, bearing in mind that there are special, detailed, regimes for natural resource and other industries that present special problems of distinguishing between capital and revenue because of the long duration of their business cycles.[21] Whichever way one looks at it, the way that the New Zealand Income Tax Act treats the capital/revenue distinction offers little comfort to people who support a drafting style that relies on broad principles rather than detailed rules.

Modern principles and purpose drafting

A much more recent example of drafting using principles rather than rules is found in the New Zealand accruals regime, a regime that was enacted in 1986 and that in principle applies to all "financial arrangements". "Financial arrangement" is defined very widely, to include not only all sorts of loans and bonds, but credit sales, swaps, assignments of income, options, futures and foreign exchange contracts, and any arrangement whereby one person receives a benefit now in consideration for another person receiving a benefit at some other time.

[20] See, *e.g. Rangatira Ltd v. Commissioner of Inland Revenue* (1996) 17 N.Z.T.C. 12 at 727, PC, and cases cited there and before the Court of Appeal 17 N.Z.T.C. 12 at 182. It is fair to say that the Privy Council's judgment in the *Rangatira* case was in favour of the taxpayer and is more or less neutral on the question of whether the scope of what is classed as revenue is increasing.

[21] See, *e.g.* Income Tax Act 1994, subparts CJ, DK, DL, DM, and DN (minerals, timber, films, petroleum mining), subparts EJ, EK, and EL (farming) and subpart CM (life insurance).

[1998] B.T.R.: No.2 © SWEET & MAXWELL AND CONTRIBUTORS 1998

Section EH1 of the Income Taxes Act is the relevant charging provision. It simply provides that all "financial arrangements" must be brought to account on a yield to maturity basis. The rest of subpart EH contains a number of exceptions and modifications to this rule. For example, some taxpayers are permitted to spread income on a straight line, rather than yield to maturity, basis. Considering the great breadth of coverage of the accruals regime, the subpart is relatively concise at 16 pages of legislation, most of it explaining exceptions to the basic yield to maturity principle.

On the face of it, subpart EH is a successful effort to legislate in reference to principle, and to eschew detail. In fact, there is a good deal more involved. The reason for the conciseness of subpart EH is not so much because that is all the legislation that was felt to be needed, but because the range of legislation required is so extensive and detailed that people concluded that it could not be fitted into even a traditional detailed-rule income tax statute. Instead, section 90 of the Tax Administration Act 1994 empowers the Commissioner of Inland Revenue to issue determinations that set out how the accrual rules will apply in different circumstances. This power has been and continues to be used extensively, as the Commissioner continues to promulgate rules to cover a great variety of circumstances.

The accruals regime attracts a good deal of criticism from people who find it hard to understand or to apply. But it is hard to see how the effect of the regime could be achieved with any other form of legislation. Legislation by reference to principles only would have led to even greater uncertainty, as taxpayers waited for the courts to rule on one case after another. A detailed matrix of rules against a background of a single general principle, the yield to maturity calculation, is at least a theoretically robust solution.

Controlled foreign company legislation

In further examining the question, principle or detail, consider controlled foreign company legislation, typically some of the most complex legislation in any income tax regime. The writer suggests that an effective controlled foreign company regime could not be drafted in terms of principle. One reason is that the most complex parts of controlled foreign company regimes tend to be dictated by administrative and taxpayer convenience, rather than by principle.

The United Kingdom, Australian, and most other controlled foreign company regimes are based on a principle of frustrating tax avoidance. Unusually, the New Zealand regime, found in subparts CG and MF of the Act, is based on achieving neutrality and a comprehensive tax base. But whatever their underlying principle, all controlled foreign company regimes are subject to significant exceptions.

Broadly speaking, legislatures engraft exceptions onto controlled foreign company regimes where it seems that the form or source of the income in question suggests that the arrangement is not avoidance-driven; where administrative or compliance costs appear likely to outweigh tax collected; and where tax at source appears likely to be high enough to make it not worthwhile for the country of residence also to tax the same income.

The United Kingdom regime contains a *de minimis* exception, and exceptions in respect of active businesses, publicly quoted companies, and companies that manage to demonstrate that they are not actuated by avoidance motives.[22] All these exceptions are

[22] For a short description of the United Kingdom regime see J. Prebble, *The Taxation of Controlled Foreign Corporations*, (Institute of Policy Studies, Wellington, 1987) pp. 33 *et seq.*

[1998] B.T.R.: No.2 © SWEET & MAXWELL AND CONTRIBUTORS 1998

carefully defined by means of detailed rules. The New Zealand regime contains an exception for companies in named jurisdictions that are thought to have effective tax rates comparable to or higher than New Zealand's, with provision for counter-exceptions when the companies in question exploit identified tax preferences. The New Zealand regime contains detailed rules for such things as determining ownership participation at any particular time, and other rules that allow taxpayers to opt for simplified calculations.[23]

It is hard to see how regimes drafted in terms of principles could achieve a result at all similar. In the United Kingdom, the first three exceptions to the controlled foreign company regime could be omitted, leaving only the exception of no tax avoidance motive, a kind of principle rather than a rule. But that would leave a great deal of uncertainty that is cleared away by the more specific *de minimis*, active business, and public quotation rules.

The New Zealand regime balances an attempt to make the tax base comprehensively watertight and an attempt to avoid generating undue compliance and administrative costs. Faced with these conflicting principles, how could a court decide where to draw the line? How would taxpayers decide which side of the line their companies fell? It is hard to see a practical alternative to bright-line rules.[23a]

Conflicting principles and conflicting rules

Theoretically, cases where two principles conflict are easier to resolve than cases where two rules conflict, because, logically, conflicting principles can exist side by side, but if rules conflict the court must decide which takes precedence. However, tax law does not lend itself to the resolution of conflict between principles any more than it lends itself to the resolution of conflict between rules. Take the case of *Budget Rent A Car Ltd v. Commissioner of Inland Revenue*.[24] In this case, a large debt owing to the taxpayer went bad. In a later year, the taxpayer wrote the debt off and claimed it as a deduction, which the taxpayer was entitled to do under the terms of what is now section DJ1 of the Act. In the meantime, the membership of the company had changed, with the result that the beneficiaries of the deduction were different persons from the owners of the taxpayer company when the debt was incurred and went bad.

The result of the procedure just outlined was to circumvent the shareholder continuity rules in the loss carry forward provisions that apply to companies by virtue of subpart IF of the Act. Formally, the company did not sustain a loss until it wrote the debt off. Thus, the new shareholders, who were the people who wanted to take advantage of the deduction, were also the people who were shareholders at the time that the loss was formally incurred. Consequently, there was no formal breach of ownership continuity. In substance, on the other hand, the company sustained the loss when the debt went bad, when the company was owned by different people.

Tompkins J. held that the bad debt rules applied, and allowed the deduction. This decision was probably inevitable, in that the bad debt rules applied literally to the taxpayer's actions, whereas the loss carry forward rules applied only in spirit.

Would the decision have been different if the rules had been stated as principles? The

[23] For a discussion of some of the simplified calculation rules see *ibid.* para. 3.6.

[23a] A bright-line rule is a legal rule that is precise and that gives precise answers to questions. For example, the rule that says that a document must have two witnesses if it is to be valid as a will is a bright-line rule. We know that a document with only one witness cannot be a will.

[24] [1995] 3 N.Z.L.R. 90, Tompkins J.

answer would depend on how generalised was the relevant principle. Suppose it had been, "Taxpayers will be assessed according to principles of economic equivalence" the decision might well have been the same, on the basis that economic principle suggests that taxpayers should be permitted to have the benefit of their losses, if necessary by dealing in them.

Suppose, on the other hand, there were two statements of principle, one saying, "Companies shall not carry losses forward in the absence of continuity of a substantial fraction of ownership" and the other saying "Bad debts that are trade debts shall be deductible when they are written off." In this event, people reading the legislation would be no better informed. The principle about company losses reflects a somewhat emotive policy against pejoratively-described "loss trafficking", and an attempt to frustrate aggressive tax planning. Somewhat unusually, the principle about bad debts allows taxpayers themselves to decide when this kind of loss will be taken into account for tax purposes. It is hard to see how a court could make a logical choice between these principles. The only obvious solution is a detailed rule that says which principle must prevail. This same solution might have helped Tompkins J. in the *Budget Rent A Car* case itself, had someone thought to draft such a rule in the first place.

What role is there for principles?

John Avery Jones advocates a two-level solution:

> At the top are more abstract principles corresponding to those found in the European Community Treaty applying generally, and the more specific principles found in the preamble to Community legislation. The second level is the legislation itself, which is interpreted with the aid of the higher-level principles as well as the explanatory memoranda. There is no third level of detailed legislation, but, of course, below the legislation is Revenue practice which is now published.[25]

With respect, this proposal must be evaluated by seeing how it might apply in particular cases. This article has examined a number of cases. In many of them the law must choose between, or must compromise, two conflicting principles. It is hard to see how the drafting pattern that Mr Avery Jones proposes would cope with this conflict.

A possible answer is: refer the conflict to a higher principle, stated in terms of greater abstraction. One would need examples to be persuaded. To the same point, does income tax law in fact have such higher principles, that can be stated in terms sufficiently specific to be useful in circumstances of this kind? The answer is probably no. Income tax law is not based in sound, *a priori* principle. Instead it is a compromise between on the one hand a desire to assess gains across a neutral, comprehensive tax base, and on the other hand all sorts of political and administrative forces: the need to nominate a geographical source for income; the need to divide income into portions delimited by time; the difficulty of valuing and taxing benefits that taxpayers enjoy from their own labour or assets; the emotional and political reaction to capital gains taxes in some jurisdictions, and so on.

The writer has argued that these factors make income tax law ectopic. It is disconnected from the facts to which it relates, and it is not always logical. These characteristics are exaggerated in marginal and difficult cases. The result is an absence of principle.

[25] J. Avery Jones, "Tax Law: Rules or Principles" (1996) 17 *Fiscal Studies* 63 at 79. No doubt, in countries like New Zealand that have provisions for the issue of binding rulings, rulings have a place in John Avery Jones's hierarchy similar to that of United Kingdom Revenue practice.

[1998] B.T.R.: No.2 © SWEET & MAXWELL AND CONTRIBUTORS 1998

People sometimes argue, "If drafters can manage briefer tax legislation in other jurisdictions, why not in the Anglophone common law countries? If value added tax law can be drafted briefly, why not income tax law?"[26]

As to value added tax law, the answer is that value added tax is not ectopic. The crucial triggering element for value added tax is the transaction, a fact that occurs, that exists, and that can be identified. Accordingly, it is possible for value added tax rules to have a relationship with their subject matter that is not merely closer than the relationship between income tax law and its subject matter, but that is different in kind.

Secondly, as to tax laws in other jurisdictions, it is notable that when subordinate legislation is taken into account United States income tax rules are just as detailed and intractable as those of the United Kingdom, Australia, and New Zealand. This position is something of an exception for the United States, which generally adopts a briefer, more conceptual drafting style than other common law countries.[27] Mr Avery Jones does not know the reason for this example of exceptional complexity in American statute law.[28] It is possible that the answer lies in the intractable nature of the subject matter.

As to how European countries manage with briefer tax legislation than Anglophone countries an answer does not so readily suggest itself. One possibility is that European countries tend to rely more on accounting practice than do the common law countries. That approach would certainly make tax compliance easier, but would not necessarily increase the tax take.

Another possibility, put forward tentatively, is that European tax systems have not yet reached the levels of sophistication of Anglophone systems, and tolerate more breaches of neutrality in their tax bases. This chauvinistic hypothesis is based more on a number of informal conversations than on exhaustive empirical study, and would need a major research project to be proved or disproved. However, one should not assume that when a European country succeeds in constructing a tax regime while using many fewer words than an Anglophone country the European country has constructed a regime that is comparably robust.

[26] *cf.* J. Avery Jones (1996) 17 *Fiscal Studies* 63 at p. 75.
[27] *ibid.* p. 74.
[28] *ibid.* p. 76.

[1998] B.T.R.: No.2 © SWEET & MAXWELL AND CONTRIBUTORS 1998

Appeal Cases

BEFORE

THE HOUSE OF LORDS

(ENGLISH—IRISH—AND SCOTTISH)

AND

THE JUDICIAL COMMITTEE

OF

HIS MAJESTY'S MOST HONOURABLE

PRIVY COUNCIL.

[HOUSE OF LORDS.]

H. L. (E.)*
1935
May 7.

THE COMMISSIONERS OF INLAND REVENUE . . . APPELLANTS;

AND

HIS GRACE THE DUKE OF WESTMINSTER . . . RESPONDENT.

Revenue—Income Tax—Surtax—Annual Payments under a Deed—Salaries and Wages—Provision made for servant—Covenant—Acknowledgment in writing—Income Tax Act, 1918 (8 & 9 *Geo.* 5, *c.* 40), *Sch. D, Case III., Rule* 1; *All Schedules Rules, Rule* 19.

By a deed made in August, 1930, the respondent covenanted to pay A., a gardener in his employment, a yearly sum of 98*l.* 16*s.* by weekly payments of 1*l.* 18*s.* for a period of seven years or during the joint lives of the parties, and it was agreed that the payments were without prejudice to the remuneration to which A. should be entitled for services, if any, thereafter rendered.

---

* *Present:* LORD ATKIN, LORD TOMLIN, LORD RUSSELL of KILLOWEN, LORD MACMILLAN, and LORD WRIGHT.

H. L. (E.)
1935
INLAND REVENUE COMMISSIONERS v. WESTMINSTER (DUKE).

Before the deed was executed the respondent's solicitors on his instructions wrote to A. a letter the material parts of which were as follows: "On the 6th inst. we read over with you a deed of covenant which the Duke of Westminster has signed in your favour. . . . . We explained that there is nothing in the deed to prevent your being entitled to and claiming full remuneration for such further work as you may do, though it is expected that in practice you will be content with the provision which is being legally made for you for so long as the deed takes effect with the addition of such sum, if any, as may be necessary to bring the total periodical payment while you are still in the Duke's service up to the amount of the salary or wages which you have lately been receiving. You said that you accepted this arrangement, and you accordingly executed the deed. . . . . If you are still quite satisfied, we propose to insert the 6th inst. as the date of the deed, and we shall be obliged by your signing the acknowledgment at the foot of this letter and returning it to us." A. signed the acknowledgment accepting the provision made for him and agreeing to the deed being dated and treated as delivered and binding on the parties thereto. The acknowledgment was stamped with a sixpenny stamp:—

Held, by Lord Tomlin, Lord Russell of Killowen, Lord Macmillan, and Lord Wright (Lord Atkin dissenting), that sums paid yearly under the above mentioned documents were annual payments within Sch. D, Case III., Rule 1, and Rule 19, s. 1, of the Rules applicable to Schedules A, B, C, D, and E of the Income Tax Act, 1918, and were not payments of salary or wages; and consequently that the respondent, being entitled to deduct tax from the payments, was entitled to deduct the payments themselves in arriving at his total income for the purposes of surtax.

Order of the Court of Appeal affirmed.

Apart, however, from the question of contract with which I have dealt, it is said that in revenue cases there is a doctrine that the Court may ignore the legal position and regard what is called "the substance of the matter," and that here the substance of the matter is that the annuitant was serving the Duke for something equal to his former salary or wages, and that therefore, while he is so serving, the annuity must be treated as salary or wages. This supposed doctrine (upon which the Commissioners apparently acted) seems to rest for its support upon a misunderstanding of language used in some earlier cases. The sooner this misunderstanding is dispelled, and the supposed doctrine given its quietus, the better it will be for all concerned, for the doctrine seems to involve substituting "the incertain and crooked cord of discretion" for "the golden and streight metwand of the law." (1) Every man is entitled if he can to order his affairs so as that the tax attaching under the appropriate Acts is less than it otherwise would be. If he succeeds in ordering them so as to secure this result, then, however unappreciative the Commissioners of Inland Revenue or his fellow taxpayers may be of his ingenuity, he cannot be compelled to pay an

(1) 4 Inst. 41.

C 2

H. L. (E.)
1935
INLAND REVENUE COMMISSIONERS
v.
WESTMINSTER (DUKE).
Lord Tomlin.

increased tax. This so-called doctrine of "the substance" seems to me to be nothing more than an attempt to make a man pay notwithstanding that he has so ordered his affairs that the amount of tax sought from him is not legally claimable.

The principal passages relied upon are from opinions of Lord Herschell and Lord Halsbury in your Lordships' House. Lord Herschell L.C. in *Helby* v. *Matthews* (1) observed: "It is said that the substance of the transaction evidenced by the agreement must be looked at, and not its mere words. I quite agree;" but he went on to explain that the substance must be ascertained by a consideration of the rights and obligations of the parties to be derived from a consideration of the whole of the agreement. In short Lord Herschell was saying that the substance of a transaction embodied in a written instrument is to be found by construing the document as a whole.

Support has also been sought by the appellants from the language of Lord Halsbury L.C. in *Secretary of State in Council of India* v. *Scoble.* (2) There Lord Halsbury said: "Still, looking at the whole nature and substance of the transaction (and it is agreed on all sides that we must look at the nature of the transaction and not be bound by the mere use of the words), this is not the case of a purchase of an annuity." Here again Lord Halsbury is only giving utterance to the indisputable rule that the surrounding circumstances must be regarded in construing a document.

Neither of these passages in my opinion affords the appellants any support or has any application to the present case. The matter was put accurately by my noble and learned friend Lord Warrington of Clyffe when as Warrington L.J. in *In re Hinckes, Dashwood* v. *Hinckes* (3) he used these words: "It is said we must go behind the form and look at the substance . . . but, in order to ascertain the substance, I must look at the legal effect of the bargain which the parties have entered into." So here the substance is that which results from the legal rights and obligations of the parties ascertained upon

(1) [1895] A. C. 471, 475. (2) [1903] A. C. 299, 302.
(3) [1921] 1 Ch. 475, 489.

ordinary legal principles, and, having regard to what I have already said, the conclusion must be that each annuitant is entitled to an annuity which as between himself and the payer is liable to deduction of income tax by the payer, and which the payer is entitled to treat as a deduction from his total income for surtax purposes.

There may, of course, be cases where documents are not bona fide nor intended to be acted upon, but are only used as a cloak to conceal a different transaction. No such case is made or even suggested here. The deeds of covenant are admittedly bona fide and have been given their proper legal operation. They cannot be ignored or treated as operating in some different way because as a result less duty is payable than would have been the case if some other arrangement (called for the purpose of the appellants' argument "the substance") had been made.

I find myself, therefore, in regard to the annuities other than that of Blow, unable to take the same view as the noble and learned Lord upon the Woolsack.

In my opinion in regard to all the annuities the appeal fails and ought to be dismissed with costs.

LORD RUSSELL OF KILLOWEN. My Lords, I would dismiss this appeal.

It is conceded that the deeds are genuine deeds, i.e., that they were intended to create and do create a legal liability on the Duke to pay in weekly payments the annual sum specified in each deed, whether or not any service is being rendered to the Duke by the covenantee. Further, it is conceded that the sums specified in the deeds were paid to the covenantees under the deeds.

The question for our decision is whether those sums so paid constitute part of the Duke's income for the purpose of computing his liability for surtax in the particular years in question.

I need not consider in detail the various statutory provisions which are relevant to the consideration of this matter. The result may for the purposes of this case be summarized thus:

H. L. (E.)
1935
INLAND REVENUE COMMISSIONERS v. WESTMINSTER (DUKE).
Lord Russell of Killowen.

If the payment of these sums is payment of salary or wages within Sch. E (1), from which tax is not deductible by the Duke, then he is not entitled to exclude the amounts paid in ascertaining his total income for surtax purposes, but if the payment is an annual payment within Sch. D, from which tax is deductible by the Duke, then he is entitled to exclude the amounts paid in ascertaining such total income.

There can I think be no doubt that if the deeds stood alone the payments are annual payments within Sch. D. Indeed, this is not I think disputed. It is, however, argued that certain letters written by the Duke's solicitor to the covenantees and certain acknowledgments signed by the covenantees at the foot of those letters, effect a complete change in the situation, and turn the payments made under the deeds into payments of salary and wages within Sch. E.

I will consider this suggestion in relation to the case of Frank Allman. The argument centred round his case, and it was common ground that all the cases (with the exception of the case of Mr. Blow) stood or fell together notwithstanding any difference of wording which might exist among them.

The legal position created by Allman's deed is clear. He is entitled during the defined period to his annual sum of 98*l.* 16*s.* by weekly payments of 1*l.* 18*s.*, commencing on August 9, 1930. He is not bound to do a stroke of work in order to be entitled to payment. If he does in the future render any service to the Duke, he will be legally entitled to claim remuneration for it, over and above the payments under the deed, which are to be without prejudice to his remuneration for future services. The deed expressly so provides.

The letter to Allman states the effect of the deed, but says that it is expected that in practice he will be content with the legal provision made by the deed "with the addition of such sum (if any) as may be necessary to bring the total periodical payment while you are still in the Duke's service up to the amount of the salary or wages which you have lately been receiving." That is an expression of hope or anticipation, that the covenantee will not enforce his legal right to

(1) See Finance Act, 1922 (12 & 13 Geo. 5, c. 17), s. 18, sub-ss. 1, 2.

remuneration for future services beyond a certain amount. The letter states that the covenantee had "accepted this arrangement" and asks him to sign an acknowledgment in a form already written out at the foot of the letter. The arrangement said to have been accepted can be nothing more than what the letter states—namely, the execution of a deed which was to be binding and in full force, coupled with an expectation on the part of the Duke that the covenantee's legal right to full remuneration for future services would not be enforced. There is no evidence of any other arrangement. Acceptance of that arrangement cannot turn the expectation into an enforceable legal right. The acknowledgment signed by the covenantee is in strictly limited terms. It accepts the provision made by the deed; it in no way admits or suggests that the deed has to any extent been qualified by the letter. My Lords, for myself I can find nothing in the letter and acknowledgment which constitutes or resembles a contract, notwithstanding the fact that the names of the solicitors were written across an adhesive stamp. There is an expression of a hope or anticipation or expectation that the covenantee will pursue a certain line of conduct, but he nowhere binds himself to do so, nor indeed is he even asked to do so. In my opinion the letter has no operation at all, and has no effect upon the legal rights and liabilities of the parties created by the deed.

But if I am wrong in this view, and some contract dehors the deed was brought into existence by means of the letter and acknowledgment, it can be no more than a contract by Allman that his remuneration for future services shall not be full remuneration but only the additional sum referred to in the letter. I can see no grounds for extracting from the language used a contract that the remuneration for future services shall, despite the deed, be the sums payable under the deed in respect of past services plus the additional sum mentioned in the letter. I can find no possible justification for this. A suggestion was made that such a contract can be found by reason of the presence in the letter of the words "to bring the total periodical payment up to the amount of the salary

H. L. (E.)
1935
INLAND REVENUE COMMISSIONERS v. WESTMINSTER (DUKE).
Lord Russell of Killowen.

which you were receiving previously to the deed of covenant." I fail to see how these words can bear this strain. Indeed, to me they seem to point in the opposite direction. They recognize that full remuneration for future services will not be paid, and that the total periodical payment will be composed in part of salary and in part of something which is not salary at all.

If the true view is that (contrary to my opinion) a contract has been made to accept less than full remuneration for future services, the position is still the same—namely, that the legal rights and liabilities of the parties created by the deed remain unqualified and unaffected.

The result is that payments, the liability for which arises only under the deed, are not and cannot be said to be payments of salary or wages within Sch. E. They cannot with any regard to the true legal position be said to arise from an employment. They are, and can only be said to be, annual payments within Sch. D. Tax was deductible on payment; they are income of the recipient, and are accordingly not part of the Duke's total income for the purpose of calculating his liability for surtax.

The Commissioners and Finlay J. took the opposite view on the ground that (as they said) looking at the substance of the thing the payments were payments of wages. This simply means that the true legal position is disregarded, and a different legal right and liability substituted in the place of the legal right and liability which the parties have created. I confess that I view with disfavour the doctrine that in taxation cases the subject is to be taxed if, in accordance with a Court's view of what it considers the substance of the transaction, the Court thinks that the case falls within the contemplation or spirit of the statute. The subject is not taxable by inference or by analogy, but only by the plain words of a statute applicable to the facts and circumstances of his case. As Lord Cairns said many years ago in *Partington* v. *Attorney-General* (1): "As I understand the principle of all fiscal legislation it is this: If the person sought to be taxed comes within the letter of the law he must be taxed, however great the hardship may appear to the

(1) (1869) L. R. 4 H. L. 100, 122.

judicial mind to be. On the other hand, if the Crown, seeking to recover the tax, cannot bring the subject within the letter of the law, the subject is free, however apparently within the spirit of the law the case might otherwise appear to be." If all that is meant by the doctrine is that having once ascertained the legal rights of the parties you may disregard mere nomenclature and decide the question of taxability or non-taxability in accordance with the legal rights, well and good. That is what this House did in the case of *Secretary of State in Council of India* v. *Scoble* (1); that and no more. If, on the other hand, the doctrine means that you may brush aside deeds, disregard the legal rights and liabilities arising under a contract between parties, and decide the question of taxability or non-taxability upon the footing of the rights and liabilities of the parties being different from what in law they are, then I entirely dissent from such a doctrine.

The substance of the transaction between Allman and the Duke is in my opinion to be found and to be found only by ascertaining their respective rights and liabilities under the deed, the legal effect of which is what I have already stated.

The case of Mr. Blow's deed, which is uncomplicated by any letter, is necessarily decided, in my view, in the same way as Allman's case.

For these reasons I am of opinion that the order of the Court of Appeal was right and ought to be affirmed.

A HIGH COURT OF JUSTICE (CHANCERY DIVISION)—8, 9, 10, AND 24 JULY 1997

COURT OF APPEAL—5, 6 AND 23 OCTOBER 1998

LAW LIBRARY
- 8 AUG 2001

B

HOUSE OF LORDS—13 AND 14 NOVEMBER AND [illegible]

VICTORIA UNIVERSITY

C

MacNiven (H.M. Inspector of Taxes) v. Westmoreland Investments Ltd.(2)

D *Corporation tax—Charges on income—Interest—Investment company—Carry forward—Ramsay principle—Property investment company winding down business—Accumulated, loan interest owed to ultimate parent company—Scheme under which loans made and advances used to pay off accrued interest and loan principal—Company left owning only gilt-edged stock and money on deposit, but another property acquired to preserve status of investment* E *company—Inspector who considered first further loan accepting interest as a charge on income and agreeing that assessment under appeal be reduced to nil profits—Taxes Management Act* 1970, *s* 54, *Income and Corporation Taxes Act* 1988, *ss* 75, 130, 338 *and* 787.

F By 1983 WIL, an investment company, was, via an intermediary company, wholly owned by ESPS, a pension scheme, which was an exempt approved scheme for tax purposes. Following a report in 1980 an orderly wind down of WIL's business had been set in train. That process included the making of loan agreements in respect of £9.5m at 17 per cent. interest, to replace loan stock issued by WIL in 1976 and redeemable in 1982. Further G loans were made by ESPS to WIL over the next six years. Some interest was paid, usually, but not always, with the help of a further loan. All the properties owned by WIL were sold by March 1985, except for one which was sold by March 1988. By March 1987 WIL owed ESPS some £73m, which included some £42m unpaid interest.

H In 1988 WIL's directors received a report from accountants on the future of WIL. The advice was that WIL would be of value to a purchaser company which wished to reduce its tax liabilities, but only if WIL crystallised tax losses by paying its interest liabilities. It was suggested that, in order to ascertain whether the Revenue would accept or would challenge the concept, ESPS should initially lend £20m to WIL to enable WIL to pay some of the I accrued interest liabilities. That loan was made, and WIL thereupon paid off £20,220,000 referable to loans made in 1980 (£14,760,600 being paid to ESPS and the balance by way of tax to the Revenue from which ESPS, as an exempt scheme, would recover it). Initially the Inspector of Taxes suggested

---

(1) [2001] UKHL 6.

(2) Reported (ChD) [1997] STC 1103; (CA) [1998] STC 1131; (HL) [2001] 1 All ER 865; [2001] 2 WLR 337; [2001] STC 237.

that relief should be denied under s 787 Income and Corporation Taxes Act 1988 on the ground that the sole or main benefit from the transaction was a reduction in tax liability but later he accepted the computation and agreed that the appeal against an assessment for the accounting period ended 31 March 1988 should be determined under s 54 Taxes Management Act 1970 in the sum of nil profits chargeable to corporation tax.

In due course, in 1989 and 1990, ESPS made further loans, enabling WIL to pay further amounts of some £103m in total (some £50m in principal, and some £53m, including the recoverable tax deductions, in interest) to ESPS.

During the, period from 1988 to 1990 WIL owned some gilt-edged stock and some money on deposit, but owned no properties except for one which it bought in early 1990 from ESPS, at its open market value, and sold later in 1990. That purchase followed advice that steps should be taken to preserve WIL's status as an investment company.

In December 1990 an outside company CSB, acquired WIL's share capital for £1 and its debts (together with some debts owed by associated companies) for some £2m. Later CSB transferred income-producing assets to WIL in consideration of an issue of new shares.

WIL appealed against assessments to corporation tax for six accounting periods ended on various dates between 1988 and 1992, which were raised on the basis that the payments of interest made by WIL in 1988 to 1990 were not allowable charges on income. The Crown advanced several contentions in support, but it was common ground that the £220,000 paid by WIL along with the £20m, would be an allowable charge unless the Crown succeeded in the contention that WIL had not been an investment company from 1988 to 1990. The Special Commissioners[1] allowed WIL's appeal, rejecting all of the Crown's several contentions, and finding it unnecessary, to decide a contention on behalf of WIL that, by the agreement under s 54 of the 1970 Act, the Inspector had agreed not only the computation for the period under review but also amounts which could be carried forward. The Crown appealed on five grounds, of which the Judge found it necessary to consider only two. WIL maintained its contention based on the agreement under s 54 of the 1970 Act.

The Chancery Division[2] held, allowing the Crown's appeal, except as to the payment of £220,000, that:

(1) WIL was an investment company throughout the period, including 1988 to 1990; there had never been any intention to change its previous character as an investment company and, on the contrary, steps had been taken between 1988 and 1990 to preserve that character; at the very least it had carried out "a residual function", as an investment company; the Commissioners had been correct, and had certainly been entitled, to look at WIL's activities over a longer period than the three years under review, and to conclude that there had been no "definite change" in the type of business;

(2) the *Ramsay* principle applied, so that the payments of interest to ESPS in 1988 to 1990, which were made wholly out of money borrowed from

(1) Pages 7–22 *post*. (2) Pages 22–39 *post*.

A ESPS were not payments of interest for the purpose of s 338 Income and Corporation Taxes Act 1988 and were, therefore, not allowable charges on income; the arrangements, following the advice given in 1988, for replacements of the old loans by new ones were a preordained series of transactions; the transfer of funds from ESPS to WIL, and back again, were steps inserted for no commercial purpose apart from the avoidance of B liability to tax, that is the conversion of a notional interest obligation which had no tax significance into one which could be used to reduce future tax liabilities; those circular transfers were to be ignored; the payments were to be treated for tax purposes as never having happened; the fact that some of the money returned to ESPS via a diversionary route through the Revenue did not affect the matter, that diversion not being a purpose in its own right C but merely an incidental effect of the arrangement;

(3) the agreement under s 54 of the 1970 Act did not extend to amounts to be carried forward: while the reasoning of the Inspector who made the agreement had extended to the amounts intended to be carried forward, that did not mean that the agreement was binding on the Revenue; that depended D on the terms of s 54; s 54 gave an agreement the like effect as a decision of the Commissioners; an appeal to the Commissioners would have been solely against an assessment in respect of the particular year, so that the Commissioners would have had no jurisdiction to determine tax liabilities for future years;

E *Tod* v. *South Essex Motors (Basildon) Ltd.* [1988] STC 392: 60 TC 598 and *Barnett* v. *Brabyn* [1996] STC 716: 69 TC 133 followed.

WIL appealed. By a Respondent's notice, the Crown relied on several arguments other than that the *Ramsay* principle applied.

F The Court of Appeal(1) held, allowing WIL's appeal, that:—

(1) the *Ramsay* principle did not apply, because:

(a) this was not a case where an artificially contrived concatenation of G transactions had been devised, for example, to convert an unallowable loss into an allowable loss or the receivability of a dividend into the receipt of capital; WIL had had a genuine accrued interest liability under genuine loans, and it had been in WIL's commercial interest to refinance its indebtedness, because of the burdensome rates of interest chargeable under those loans; s 338 of the 1988 Act makes clear that the debtor is entitled to H treat interest as a charge on income, once it is paid; although it was ESPS, acting in its own interests, which had funded its indirect subsidiary with a view to increasing the saleability of the shareholding in WIL to a purchaser, it had also been in WIL's interests that it should crystallise that liability as a charge on income to be carried forward as excess management expenses capable, for tax purposes, of being set against future profits which it might I make; from the viewpoint of WIL, it had been possible that assets might have been injected into WIL so that it would generate profits and might have been able to pay off its debts; creditors and debtors with accrued interest liabilities not infrequently renegotiated those liabilities; they might do so by simply varying the terms of the debts; they might also do so by a further increased loan coupled with the repayment of the accrued interest; frequently

---

(1) Pages 39–55 *post*.

a debtor refinanced his debt including an accrued interest liability by borrowing an increased sum so as to discharge that debt; it was not possible as a matter of statutory construction to deny to the parties the consequences of the exercise of their right to adopt whatever course was more advantageous to them to crystallise a genuine tax loss, even if there was a tax purpose influencing their choice; and

(b) the case involved tax mitigation rather than tax avoidance; it could not be accepted that the insolvent WIL had not genuinely suffered the economic consequences which Parliament intended to be suffered by a payer of interest; the fact that WIL had borrowed from ESPS the money used to pay the interest did not mean that the payment of interest was not genuine as the loan was found by the Commissioners to be; WIL could not be sued for the interest paid once ESPS received it; Parliament intended yearly interest to be allowed as a deduction in computing a company's profits liable to corporation tax and introduced a symmetry of sorts between payer and payee; by s 338 it allowed interest on a bank loan to be deducted as it accrued in the bank's books because the bank would be taxed on it on the same basis; on loans from lenders other than banks, it allowed interest to be deductible when paid, because the ordinary lender would be taxed on the interest which it received; the abnormal feature of the present case was that the lender was an exempt approved scheme not liable to tax on the interest paid by WIL, but it would be surprising if that feature with the tax advantage provided to ESPS by Parliament gave the Revenue the right to attack what was done by the application of the *Ramsay* principle;

(2) the grounds, other than the *Ramsay* principle, on which the Crown relied, failed because—

(a) the payments of interest had ultimately been borne by WIL; the reference in s 338(5)(a) of the 1988 Act to payments not ultimately borne by a company was to situations where a third company made a grant or contribution to the company making payment, such as where the company had a joint and several liability with another from whom the company obtained a contribution; that was the more natural interpretation than one which required an assessment to be made of the commerciality of the transaction financing the payment;

(b) the Commissioners had been correct in concluding as a fact that the payments of interest were wholly and exclusively made for the purpose of WIL's business, and not for ESPS's purpose;

Hyett v. *Lennard* [1940] 2 KB 180: 23 TC 346 considered;

and (c) s 787(1) of the 1988 Act, which provided that "relief shall not be given to any person under any provision of the Tax Acts in respect of any payment of interest if a scheme has been effected or arrangements made (whether before or after the time when the payment is made) such that the sole or main benefit that might be expected to accrue to that person from the transaction under which the interest is paid was the obtaining of a reduction in tax liability by means of such relief", did not apply; the transaction under which the interest was paid was the loan agreement, not the transaction which made it possible for the interest to be paid; s 787, on its natural construction, did not require an examination of the circumstances surrounding each payment of interest:

A but (3) the High Court Judge had been right, for the reasons which he gave, to reject WIL's alternative argument based on s 54 of the 1970 Act.

The Crown appealed.

B *Held*, in the House of Lords(1), dismissing the Crown's appeal that:—

(1) as to the *Ramsay* principle generally:—

(a) there is ultimately only one principle of construction, namely to ascertain what Parliament meant by using the language of the statute; all C other "principles of construction" can be no more than guides which past Judges have put forward, some more helpful or insightful than others, to assist in the task of interpretation;

(b) the *Ramsay* principle is giving statutory language such a construction D as to interpret statutory words in a commercial sense which transcends the individuality of the intermediate stages of a transaction and looks at the end result; whether the *Ramsay* principle is applicable depends upon the language and purpose of the statutory provision concerned; it is first necessary to construe the statutory language and decide whether it refers to a concept which Parliament intended to be given a commercial meaning capable of E transcending the juristic individuality of its component parts; many expressions used in tax legislation (and elsewhere) can be construed as referring to commercial concepts and the Courts are today readier to give them such a construction than they were before *Ramsay*; that is not always the case, because there are many terms in tax legislation which cannot be construed in that way, as they refer to purely legal concepts which have no F broader commercial meaning; but, where a word does have a "recognised legal meaning", the legislative context may show that it is in fact being used to refer to a broader commercial concept;

(c) if the legal position is that tax is imposed by reference to a commercial concept, then for the Courts to have regard to the business G substance of the matter is not to ignore the legal position but to give effect to it;

(d) even if a statutory expression refers to a business or economic concept, it is not possible to disregard a transaction which comes within the statutory language, construed in the correct commercial sense, simply on the H ground that it was entered into solely for tax reasons; likewise the use of business concepts like "income" and "capital" may give the taxpayer a choice of structuring a commercial transaction so as to come within one concept or the other; further, a transaction which, for the avoidance of tax, has been structured to produce, say, capital, and does produce capital in the ordinary commercial sense of that concept, cannot be "recharacterised" as producing I income; but an attempt to relabel a sum of money by a transaction which has no commercial purpose fails to transform an income receipt from one person into a capital receipt from another;

(2) as to the present case:—

---

(1) Pages 56–84 *post*.

(a) payment of a debt such as interest ordinarily means an act, such as the transfer of money, which discharges the debt, and on the facts the interest debt was indeed discharged; there is no alternative concept of payment which could apply; in particular, it had not been contended that payment must involve a negative cash flow which is not compensated by a cash flow in the opposite direction; many commercial refinancing operations discharge old debts and create new ones without any cash flow either way; further there is no apparent policy to be found in s 338 which would require a negative cash flow; the Crown's real complaint was that ESPS, as an exempt fund, was able to reclaim the tax, but that could not be remedied by giving the word "paid" a different meaning, because that word must mean the same, whatever the status of the lender; the Crown's objection to the circularity of the cash flow, combined with the fact that the transaction took place entirely for tax purposes, could not affect the conclusion, because payment was a legal concept which did not have some other commercial meaning;

(b) in respect of the Crown's alternative arguments, for the reasons given by the Court of Appeal, none of the three provisions (s 338(5)(a), s 75(3) and s 787(1)), relied upon as nullifying the effect of the payment of interest, had that effect.

Dicta of Lord Cooke of Thorndon in *Commissioners of Inland Revenue* v. *McGuckian* [1997] 1 WLR 991: 69 TC 1, and decision of Langley J. in *NMB Holdings Ltd.* v. *Secretary of State for Social Security* (14 July 2000, 73 TC 85), approved.

Sun Insurance Office v. *Clark* [1912] AC 443: 6 TC 59, *Commissioners of Inland Revenue* v. *Duke of Westminster* [1936] AC 1; 19 TC 490, *Commissioners of Inland Revenue* v. *Wesleyan and General Assurance Society* (1946) 30 TC 11, *Gilbert* v. *Commissioners of Inland Revenue* (1957) 248 F2d 399; *Southern Railway of Peru Ltd.* v. *Owen* [1957] AC 334: 36 TC 602, *W.T. Ramsay Ltd.* v. *Commissioners of Inland Revenue* [1982] AC 300: 54 TC 101, *Commissioners of Inland Revenue* v. *Burmah Oil Co. Ltd.* [1982] STC 30: 54 TC 200, *Cairns* v. *MacDiarmid* [1982] STC 226: 56 TC 556, *Furniss* v. *Dawson* [1984] AC 474: 55 TC 324, *Customs & Excise Commissioners* v. *Faith Construction Ltd.* [1989] STC 539, *Barclays Bank plc* v. *British Commonwealth Holdings plc* [1996] 1 WLR 1, and *Norglen Ltd.* v. *Reeds Rain Prudential Ltd.* [1999] 2 AC 1 considered.

Per Lord Hope of Craighead (Lord Hoffmann agreeing): an agreement made under s 54 has no wider effect upon the position of either party than that which has been provided for by the machinery in the Taxes Management Act 1970; that machinery is limited to determining conclusively the amount of tax chargeable for the year of assessment; it does not enable such determinations to be made, either on appeal or by agreement, as to the amounts of tax chargeable in future years.

Lord Hoffmann

I

My Lords,

The issue

19. The question in this appeal is whether certain payments of interest made by a property investment company named Westmoreland Investments Ltd. ("WIL") in the years 1988 to 1990 were "charges upon income" within the

A meaning of s 338 of the Income and Corporation Taxes Act 1988 and therefore allowable deductions in computing its profits or losses for the purposes of corporation tax. I speak of them as payments in the sense that there is no dispute that WIL transferred money to the lender and that its liability for interest was thereby discharged. As between the parties, the interest was paid. But the dispute between the taxpayer and the Crown is whether the interest was B "paid" within the meaning of s 338. It arises because WIL paid the interest out of money which it had been lent by the lender for the specific purpose of enabling it to pay. The interest liability was replaced by a liability for an additional capital sum. The transaction was circular; WIL borrowed capital and paid it back as interest. And the only purpose of the transaction was to produce an allowable deduction for corporation tax. The Crown says that this does not C count as a payment for the purposes of the Act. It must be disregarded under the principle, in *W.T. Ramsay Ltd.* v. *Inland Revenue Commissioners* [1982] AC 300; 54 TC 101. The main issue in this appeal is therefore the meaning, scope and applicability of that principle. The Crown also says that the taxpayer's claim is defeated by three specific anti-avoidance provisions in the Taxes Act. I shall deal with these after considering the main point.

D

The statutory provisions

20. My Lords, I set out first the relevant provisions of s 338 as it stood at the relevant time. It has since been substantially amended.

E "(1) ... in computing the corporation tax chargeable for any accounting period of a company any charges on income paid by the company in the accounting period, so far as paid out of the company's profits brought into charge to corporation tax, shall be allowed as deductions against the total profits for the period ...

(2) ... 'charges on income' means for the purposes of corporation F tax—(a) payments of any description mentioned in subsection (3) below ...

(3) Subject to subsections (4), to (6) below, the payments referred to in subsection 2(a) above are—

(a) any yearly interest ... and

G

(b) any other interest ... payable in the United Kingdom on an advance from a bank carrying on a *bona fide* banking business in the United Kingdom ...

and for the purposes of this section any interest payable by a company as mentioned in paragraph (b) above shall be treated as paid on its H being debited to the company's account in the books of the person to whom it is payable."

21. A company is therefore allowed a deduction in respect of bank interest immediately it is debited in the books of the bank. It does not matter whether the liability has been discharged or not. But other yearly interest is I deductible only when it has been paid. Why the distinction? It reflects the difference in the way in which the Crown recovers tax from the recipient of the interest. In the ordinary case of yearly interest, the person who pays must deduct the tax and account to the Revenue: see s 349(2). But this rule does not apply to banks: see s 349(3)(a). A person who pays interest to a bank does not deduct tax. The interest is part of the bank's trading income and must be brought into account in the computation of profits when it falls due

and is debited to the borrower in its books. In both cases, therefore, the provisions of s 338 and 349 synchronise the payer's right to a deduction and Crown's right to treat the interest as a taxable receipt of the payee. A

The facts

22. My Lords, the relevant facts can be briefly summarised. WIL was owned by the Electricity Supply Pension Scheme ("the scheme"), an approved superannuation scheme which is exempt from income tax. In the early 70's it used WIL as a vehicle for some very ill-advised property investments. These were financed by money lent to WIL by the scheme. After the final liquidation of its properties in 1988, WIL had virtually no assets and a huge indebtedness to the scheme. This included over £40 million arrears of interest. B C

23. It might have been thought that the scheme had no option but to allow WIL to go quietly into liquidation. But even in its moribund state, WIL was not without its attractions. There was at the time a market in companies with established tax losses. If profit-earning assets were transferred to such companies, they could avoid tax until they had exhausted the right of set-off against losses carried forward from the earlier years. People were prepared to pay for tax loss companies, which were on offer at a very substantial discount to the expected savings they could provide. The difficulty for the scheme was that although WIL's losses were only too real, they were partially represented in the company's accounts by unpaid arrears of interest. Under s 338, these sums became deductible only when paid. But WIL had no money with which to pay the interest and no assets upon which money could be raised. D E

24. The scheme therefore lent WIL the money with which to pay the interest. On 28 January 1988 it lent £20m, repayable "as and when the company is able" with interest at 2 per cent. over base rate. On the same day WIL paid the scheme £14,760,600 net of tax (representing a gross payment of £20,220,000 interest) and accounted to the Inland Revenue for £5,459,400 tax. If nothing more had happened, the Crown would no doubt have viewed matters with equanimity. Any deduction allowed to WIL, giving rise to established losses which could be set off against such profits as might be earned at some future date, would have been more than compensated by a solid and immediate payment of tax in the same amount. But the scheme was exempt from income tax and therefore entitled to reclaim the tax from the Inland Revenue. On 17 October 1989 and 3 January 1990 the exercise was repeated. The scheme made loans to WIL which it immediately used to pay arrears of interest due under the earlier loans, accounting to the Inland Revenue for tax which was then reclaimed by the scheme. F G H

25. As a result of these transactions, the scheme was able to find a purchaser for the shares and loan debts of WIL. On 20 December 1990 a development company bought the shares for a nominal sum and the indebtedness of over £100m for 2*p* in the £. The scheme realised £2m for assets which otherwise would have been worth nothing. I

The findings of the Special Commissioners.

26. The Special Commissioners made the following findings:(1)

(1) Page 18F *ante*.

A "We find that all the loans made to WIL from 1980 onwards were real loans and WIL used them for real purposes, *viz* the discharge of real earlier outstanding loans and the payment of real accrued interest, temporary investment in part and the payment of income tax in pursuance of the statutory obligation in that behalf ...

B We do not find that the interest free loans made by the scheme in 1988/89 and 1989/90 were different in character from the earlier loans ... [A]s Mr. Milne Q.C. submits on behalf of WIL, the object of the refinancing was to crystallise the actual loss by paying interest which hitherto had merely been accrued and had not been paid. There is no question but that that accrued interest was real."

C

The Commissioners therefore held that the interest had been "paid" within the meaning of s 338(1) of the Act and gave rise to an allowable deduction.

The case for the Crown

D 27. Mr. McCall Q.C., who appeared for HM Inspector, said he did not challenge the findings of the Commissioners that the loans and payments were real in the sense that the interest debt was discharged and replaced by a loan. The transactions were not a pretence. But they had no commercial purpose. They were purely for the purpose of avoiding tax and therefore fell within the *Ramsay* principle. Under that principle, they should be E disregarded. Whatever might be their legal effect as between the parties, the absence of a commercial purpose meant that they did not count as payments within the meaning of s 338.

Ramsay: a principle of construction?

F 28. Everyone agrees that *Ramsay* is a principle of construction. The House of Lords said so in *Inland Revenue Commissioners* v. *McGuckian* [1997] 1 WLR 991; 69 TC 1. But what is that principle? Mr. McCall formulated it as follows in his printed case:

"When a court is asked

G (i) to apply a statutory provision on which a taxpayer relies for the sake of establishing some tax advantage

(ii) in circumstances where the transaction said to give rise to the tax advantage is, or forms part of, some pre-ordained, circular, self-cancelling transaction

H (iii) which transaction though accepted as perfectly genuine (i.e. not impeached as a sham) was undertaken for no commercial purpose other than the obtaining of the tax advantage in question

then (unless there is something in the statutory provisions concerned to indicate that this rule should not be applied) there is a rule of I construction that the condition laid down in the statute for the obtaining of the tax advantage has not been satisfied."

29. My Lords, I am bound to say that this does not look to me like a principle of construction at all. There is ultimately only one principle of construction, namely to ascertain what Parliament meant by using the language of the statute. All other "principles of construction" can be no more than guides which past Judges have put forward, some more helpful or

insightful than others, to assist in the task of interpretation. But Mr. McCall's formulation looks like an overriding legal principle, superimposed upon the whole of revenue law without regard to the language or purpose of any particular provision, save for the possibility of rebuttal by language which can be brought within his final parenthesis. This cannot be called a principle of construction except in the sense of some paramount provision subject to which everything else must be read, like s 2(2) of the European Communities Act 1972. But the courts have no constitutional authority to impose such an overlay upon the tax legislation and, as I hope to demonstrate, they have not attempted to do so.

Ramsay: the fountainhead

30. As is well known, the *Ramsay* case [1982] AC 300; 54 TC 101 was concerned with a tax avoidance scheme designed to manufacture a capital loss to set off against a capital gain. The question before the House was whether a transaction by which the taxpayer company acquired certain shares for £185,034 and almost immediately sold them for £9,387, gave rise to a "loss accruing on a disposal of an asset" within the meaning of s 23(1) of the Finance Act 1965. Both the acquisition and sale of the shares formed part of a pre-planned series of transactions by which the alleged loss was exactly balanced by a gain which was alleged to fall within an exemption from the charge. The aggregate effect was that the taxpayer suffered no loss except the payment of a fee to the promoters of the scheme.

31. It was not disputed that the transaction included a genuine purchase of the shares for £185,034 and a genuine sale of the same shares for £9,387. The taxpayer said that that was the end of the matter. To look at the transaction as a whole would be to commit the heresy condemned by Lord Tomlin in *Inland Revenue Commissioners* v. *Duke of Westminster* [1936] AC 1, at page 19; 19 TC 490 as the "doctrine that the Court may ignore the legal position and regard what is called 'the substance of the matter'". At first, the Revenue agreed. Its attack on the scheme concentrated on whether the counterbalancing gain really was outside the charge to tax. In the House of Lords, however, Mr. Millett Q.C. argued that no loss within the meaning of the Act had accrued at all. The House accepted the argument. Lord Wilberforce said, at page 323, that while Lord Tomlin's statement was a "cardinal principle", it did not require a court to "look at a document or a transaction in blinkers". The capital gains tax was, see [1982] AC 300, at page 326:

> "... a tax on gains (or I might have added gains less losses), it is not a tax on arithmetical differences. To say that a loss (or gain) which appears to arise at one stage in an indivisible process, and which is intended to be and is cancelled out by a later stage, so that at the end of what was bought as, and planned as, a single continuous operation, *there is not such a loss (or gain) as the legislation is dealing with*, is in my opinion well and indeed essentially within the judicial function." (My emphasis).

32. My Lords, it is worth pausing at this point to examine the characteristically compressed reasoning in a little more detail. A loss which arises at one stage of an indivisible process and cancelled out at a later stage of the same process is "not such a loss as the legislation is dealing with". The tax was not imposed "on arithmetical differences". In that case, what kind of loss was the legislation dealing with? The contrast being made throughout

Lord Wilberforce's speech is between juristic or arithmetical realities on the one hand and commercial realities on the other. He is construing the words "disposal" and "loss" to refer to commercial concepts which are not necessarily confined by the categories of juristic analysis. In the *Ramsay* case [1982] AC 300; 54 TC 101, a director, or an accountant concerned to present a true and fair view of the taxpayer's dealings, would not have said that the company had entered into a transaction giving rise to a loss which happened to have been offset by a corresponding gain. There had never been any commercial possibility that the transactions would not have cancelled each other out. Therefore, notwithstanding the juristic independence of each of the stages of the circular transaction, the commercial view would have been to lump them all together, as the parties themselves intended, and describe them as a composite transaction which had no financial consequences. The innovation in the *Ramsay* case was to give the statutory concepts of "disposal" and "loss" a commercial meaning. The new principle of construction was a recognition that the statutory language was intended to refer to commercial concepts, so that in the case of a concept such as a "disposal", the court was required to take a view of the facts which transcended the juristic individuality of the various parts of a preplanned series of transactions.

Commercial concepts in tax legislation

33. There is nothing new about terms used in tax legislation (or, for that matter, any legislation) being construed as referring to business or commercial concepts which may not be capable of being held within the confines of purely juristic analysis. A good example is the term "profits or gains of the year of assessment" which forms the basis of the charge to tax under Case I of Sch D: see s 60 of the Income and Corporation Taxes Act 1988. In *Sun Insurance Office* v. *Clark* [1912] AC 443, at page 455; 6 TC 59, at page 78 Viscount Haldane said:

> "It is plain that the question of what is or is not profit or gain must primarily be one of fact, and of fact to be ascertained by the tests applied in ordinary business."

34. It is thus the statute itself which applies the tests of ordinary business. And for present purposes, the significant feature of applying a test of ordinary business is that it may require an aggregation of transactions which transcends their juristic individuality. In *Southern Railway of Peru Ltd.* v. *Owen* [1957] AC 334 the question was whether, in calculating its profits or gains for a year of assessment, a company could make a provision for severance pay contingently payable to its employees. The Revenue argued that each contract of employment had to be separately examined and no liability could be taken into account unless it had fallen due. Lord Radcliffe rejected this approach, at page 357:

> "The answer to the question what can or cannot be admitted into the annual account is not provided by any exact analysis of the legal form of the relevant obligation. In this case, as in the *Sun Insurance case* ([1912] AC 443), you get into a world of unreality if you try to solve your problem in that way, because, where you are dealing with a number of similar obligations that arise from trading, although it may be true to say of each separate one that it may never mature, it is the sum of the obligations that matters to the trader, and experience may

show that, while each remains uncertain, the aggregate can be fixed with some precision." A

35. My Lords, it seems to me that what Lord Wilberforce was doing in the *Ramsay* case [1982] AC 300; 54 TC 101 was no more (but certainly no less) than to treat the statutory words "loss" and "disposal" as referring to commercial concepts to which a juristic analysis of the transaction, treating each step as autonomous and independent, might not be determinative. What was fresh and new about *Ramsay* was the realisation that such an approach need not be confined to well recognised accounting concepts such as profit and loss but could be the appropriate construction of other taxation concepts as well. B

The American doctrine C

36. Lord Wilberforce, while cautioning against a facile transposition of American decisions on different statutes, approved the approach of Judge Learned Hand in one of his many judgments dealing with tax avoidance schemes: *Gilbert* v. *Commissioners of Inland Revenue* (1957) 248 F2d 399. Perhaps the seminal judgment was in *Helvering* v. *Gregory* 69 (1934) F2d 809, affirmed (1935) 293 US 465, which concerned a scheme of great simplicity. The taxpayer was a stockholder in a corporation which held some shares which she wished to realise without paying tax on the gains. Instead of having the corporation sell the shares directly to the buyer, she caused it to incorporate a subsidiary and exchange the shares for an allotment of shares in the subsidiary. The subsidiary was put into liquidation and distributed the shares to the stockholder as a dividend. She then sold them to the buyer. She claimed that the exchange of shares fell within the tax exemption for a "reorganization" of capital. On the other hand, the exchange was real enough to constitute a realisation of the gain, so that no further gain was realised on the distribution to her. In the Court of Appeals (Second Circuit) Judge Learned Hand said, at page 811, that the transfer to the subsidiary did not fall within the terms of the statutory exemption: D E F

> "we cannot treat as inoperative the transfer of ... shares by [A] [or] the issue of shares by [B] of its own shares ... [B] had a juristic personality ... All these steps were real, and, *their only defect was that they were not what the statute means*." (My emphasis). G

37. What, in that case, did the statute mean? The Supreme Court of the United States affirmed the decision in a single judgment delivered by Sutherland J. "Reorganization", he said, meant a reorganization of the business of a corporation, having some business purpose. An exemption from tax could not be construed as applicable to a transaction with no business purpose except to obtain the exemption from tax. H

The Duke of Westminster's case

38. In the *Ramsay* case [1982] AC 300; 54 TC 101 both Lord Wilberforce and Lord Fraser of Tullybelton, who gave the other principal speech, were careful to stress that the House was not departing from the principle in the *Duke of Westminster's* case [1936] AC 1; 19 TC 490. There has nevertheless been a good deal of discussion about how the two cases are to be reconciled. How, if the various juristically discrete acquisitions and disposals which made up the scheme were genuine, could the House collapse them into a composite self-cancelling transaction without being guilty of ignoring the legal position and looking at the substance of the matter? I

A 39. My Lords, I venture to suggest that some of the difficulty which may have been felt in reconciling the *Ramsay* case with the *Duke of Westminster's* case arises out of an ambiguity in Lord Tomlin's statement that the courts cannot ignore "the legal position" and have regard to "the substance of the matter". If "the legal position" is that the tax is imposed by reference to a legally defined concept, such as stamp duty payable on a document which B constitutes a conveyance on sale, the court cannot tax a transaction which uses no such document on the ground that it achieves the same economic effect. On the other hand, if the legal position is that tax is imposed by reference to a commercial concept, then to have regard to the business "substance" of the matter is not to ignore the legal position but to give effect to it.

C

The real world

40. The speeches in the *Ramsay* case [1982] AC 300; 54 TC 101 and subsequent cases contain numerous references to the "real" nature of the transaction and to what happens in "the real world". These expressions are D illuminating in their context, but you have to be careful about the sense in which they are being used. Otherwise you land in all kinds of unnecessary philosophical difficulties about the nature of reality and, in particular, about how a transaction can be said not to be a "sham" and yet be "disregarded" for the purpose of deciding what happened in "the real world". The point to hold onto is that something may be real for one purpose but not for E another. When people speak of something being a "real" something, they mean that it falls within some concept which they have in mind, by contrast with something else which might have been thought to do so, but does not. When an economist says that real incomes have fallen, he is not intending to contrast real incomes with imaginary incomes. The contrast is specifically between incomes which have been adjusted for inflation and those which F have not. In order to know what he means by "real", one must first identify the concept (inflation adjustment) by reference to which he is using the word.

41. Thus in saying that the transactions in the *Ramsay* case were not sham transactions, one is accepting the juristic categorisation of the G transactions as individual and discrete and saying that each of them involved no pretence. They were intended to do precisely what they purported to do. They had a legal reality. But in saying that they did not constitute a "real" disposal giving rise to a "real" loss, one is rejecting the juristic categorisation as not being necessarily determinative for the purposes of the statutory concepts of "disposal" and "loss" as properly interpreted. The contrast here H is with a commercial meaning of these concepts. And in saying that the income tax legislation was intended to operate "in the real world", one is again referring to the commercial context which should influence the construction of the concepts used by Parliament.

The Burmah case

42. There is no doubt that the *Ramsay* case [1982] AC 300; 54 TC 101 was widely regarded as some form of judicial legislation and the concerns of taxpayers about its true scope of and its relationship with the *Duke of Westminster's* case [1936] AC 1; 19 TC 490 were not set at rest by what some regarded as the proclamation of a revolutionary credo by Lord Diplock in *Inland Revenue Commissioners* v. *Burmah Oil Co. Ltd.* (1981) 54 TC 200, at pages 214–215:

> "It would be disingenuous to suggest, and dangerous on the part of those who advise on elaborate tax-avoidance schemes to assume, that *Ramsay's* case did not mark a significant change in the approach adopted by this House in its judicial role to a pre-ordained series of transactions (whether or not they include the achievement of a legitimate commercial end) into which there are inserted steps which have no commercial purpose apart from the avoidance of a liability to tax which in the absence of those particular steps would have been payable ... [T]he approach to tax avoidance schemes of this character sanctioned by *Ramsay* entitles your Lordships to ignore the intermediate circular book entries and to look at the end result ..."

43. The *Burmah* case also concerned the question of whether the company had suffered a loss for the purposes of capital gains tax. As in the *Ramsay* case, it had produced a loss by a circular series of transactions which had no business purpose. A subsidiary owed it a substantial sum which it could not repay. As a bad debt on capital account, this would not have been an allowable loss. *Burmah* therefore invested the same amount in shares in the subsidiary, which used the money to repay the debt and then went into liquidation. *Burmah* recovered nothing on its share investment and claimed that it had thereby suffered a loss. The House of Lords held that this was not a loss caused by a disposal within the meaning of the Act. The transaction left *Burmah* no worse off than it had been before and merely purported to convert a bad debt into an allowable loss.

44. My Lords, in retrospect the *Burmah* case is an entirely straightforward application of the construction which the *Ramsay* case gave to the concept of a disposal giving rise to a loss in the capital gains tax legislation, namely that it meant a loss in commercial terms and not a series of preplanned transactions which had no business purpose. From this construction it followed that, as Lord Diplock said, the House would "ignore the intermediate circular book entries and ... look at the end result". Lord Diplock would have been the first to acknowledge that his remarks should be read in context. To "ignore" the intermediate stages of the transaction and look at the end result is something which follows logically from the decision to construe "disposal" and "loss" in a commercial sense which transcends the individuality of the "book entries". It is that decision, to give the statutory language such a construction, which I would regard as "the *Ramsay* principle". But I think that there may have been a tendency to construe Lord Diplock's statement of the consequences of applying the *Ramsay* principle to the particular provisions with which the House was concerned as if it were itself a general principle, applicable to all tax legislation. Of course such a construction could also be applied to other provisions of the taxing Acts, but this would depend upon their language and purpose. At any rate, the generalising tendency which I have described seems to me the most likely explanation of the proposition which Mr. McCall has claimed to be the *Ramsay* principle in this appeal.

Furniss v. *Dawson*

45. My Lords, in *Furniss* v. *Dawson* [1984] AC 474; 55 TC 324 the *Ramsay* construction, which in the *Ramsay* case itself and the *Burmah* case had been used to interpret the concept of a disposal giving rise to a loss, was deployed for a different purpose. The difference is occasionally described by saying that whereas *Ramsay* was a circular transaction, *Furniss* was a linear

A transaction. The difference can conveniently be encapsulated in these metaphors, but I think it is more illuminating to concentrate on the question which the legislation required the House to answer. To explain what this was, it is first necessary to give a brief account of the facts. The Dawsons wanted to sell their shares in the family business to a company called Wood Bastow Holdings Ltd. But they wanted to postpone the payment of capital gains tax. B So they formed an Isle of Man company ("Greenjacket") and exchanged their shares in the company owning the business for an allotment of shares in Greenjacket. The advantage of this transaction was that by para 6 of Sch 7 to the Finance Act 1965, a disposal of shares to Greenjacket in exchange for an allotment of its shares was treated as a reorganisation of share capital and by para 4 of the same Schedule a disposal of shares forming part of a C reorganisation was not treated as a disposal for the purposes of capital gains tax. By a preplanned transaction, Greenjacket then sold the shares to Wood Bastow for cash. But the Revenue claimed that there had been no "real" disposal to Greenjacket. It was merely a preplanned stage in a disposal from the Dawsons to Wood Bastow and fell outside the exception for a reorganisation of share capital.

D

46. Thus, while the question in the *Ramsay* case had been whether there was a disposal giving rise to a loss, the question in the *Furniss* case was whether the disposal had been to one person rather than another. But the House decided that the *Ramsay* construction, involving, as I have said, a commercial characterisation of the relevant concept, could be equally applied E to the latter question. Greenjacket was merely an artificially introduced intermediate party which was never intended to own the shares for more than an instant. Commercially, therefore, the transaction was a transfer by the Dawsons to Wood Bastow in exchange for a payment to Greenjacket. In answering the statutory question: "To whom was the disposal made?" the fact that the shares were routed through Greenjacket was irrelevant.

F

47. The consequence of adopting this construction was spelled out by Lord Brightman [1984] AC 474, at page 527; 55 TC 324; in a passage which paraphrased what Lord Diplock had said in the *Burmah* case 54 TC 200 at page 401D and has since been quoted many times. He stated the conditions G under which the commercial nature of the transaction as a whole would transcend the juristic individuality of its parts:

> "First, there must be a pre-ordained series of transactions; or, if one likes, one single composite transaction. This composite transaction may or may not include the achievement of a legitimate commercial (i.e. business) end. The composite transaction does, in the instant case; it H achieved a sale of the shares in the operating companies by the Dawsons to Wood Bastow. It did not in *Ramsay*. Secondly, there must be steps inserted which have no commercial (business) *purpose* apart from the avoidance of a liability to tax—not 'no business *effect*'. If those two ingredients exist, the inserted steps are to be disregarded for fiscal purposes. The court must then look at the end result. Precisely how the I end result will be taxed will depend on the terms of the taxing statute sought to be applied."

48. My Lords, this statement is a careful and accurate summary of the effect which the *Ramsay* construction of a statutory concept has upon the way the courts will decide whether a transaction falls within that concept or not. If the statutory language is construed as referring to a commercial

concept, then it follows that steps which have no commercial purpose but which have been artificially inserted for tax purposes into a composite transaction will not affect the answer to the statutory question. When Lord Brightman said that the inserted steps are to be "disregarded for fiscal purposes", I think that he meant that they should be disregarded for the purpose of applying the relevant fiscal concept. In the *Furniss* case, this was the concept of a disposal by one person to another. For that purpose, and for that purpose only, the disposal to Greenjacket was disregarded. But that does not mean that it was treated, even for tax purposes, as if it had never happened. The payment by Wood Bastow was undoubtedly to Greenjacket and so far as this might be relevant for tax or any other purposes, it could not be disregarded.

49. For present purposes, however, the point I wish to emphasise is that Lord Brightman's formulation in the *Furniss* case, like Lord Diplock's formulation in the *Burmah* case, is not a principle of construction. It is a statement of the consequences of giving a commercial construction to a fiscal concept. Before one can apply Lord Brightman's words, it is first necessary to construe the statutory language and decide that it refers to a concept which Parliament intended to be given a commercial meaning capable of transcending the juristic individuality of its component parts. But there are many terms in tax legislation which cannot be construed in this way. They refer to purely legal concepts which have no broader commercial meaning. In such cases, the *Ramsay* principle can have no application. It is necessary to make this point because, in the first flush of victory after the *Ramsay*, *Burmah* and *Furniss* cases, there was a tendency on the part of the Inland Revenue to treat Lord Brightman's words as if they were a broad spectrum antibiotic which killed off all tax avoidance schemes, whatever the tax and whatever the relevant statutory provisions.

50. The distinction between commercial and legal concepts has also been drawn in other areas of legislation. So, for example, the term "financial assistance" in s 151 of the Companies Act 1985 has been construed as a commercial concept, involving an inquiry into the commercial realities of the transaction: see *Burton* v. *Palmer* (1980) 2 NSWLR 878, 889–890; *Charterhouse Investment Trust Ltd.* v. *Tempest Diesels Ltd.* [1986] BCLC 1. But the same is not necessarily true of other terms used in the same section, such as "indemnity". As Aldous L.J. said in *Barclays Bank Plc* v. *British & Commonwealth Holdings Plc* [1996] 1 WLR 1, at page 14:

> "It was submitted that as the words 'financial assistance' had no technical meaning and their frame of reference was the language of ordinary commerce, the word 'indemnity' should be similarly construed. The fallacy in that submission is clear. The words 'financial assistance' are not words which have any recognised legal significance whereas the word 'indemnity' does. It is used in the section as one of a number of words having a recognised legal meaning."

I would only add by way of caution that although a word may have a "recognised legal meaning", the legislative context may show that it is in fact being used to refer to a broader commercial concept.

Inland Revenue Commissioners v. *McGuckian [1997] 1 WLR 991; 69 TC 1*

51. In the *McGuckian* case a Republic of Ireland company called Ballinamore had substantial distributable reserves. The shareholders, Mr.

A and Mrs. McGuckian, wanted to receive this money but not to pay income tax on the dividend. So they entered into a scheme by which they first transferred their shares to an offshore trustee called Shurltrust. By a series of preplanned transactions, it then assigned the right to receive the dividend to a UK company called Mallardchoice in consideration of the payment of a sum equal to 99 per cent. of the expected dividend. Ballinamore then B declared the dividend and paid it to Mallardchoice, which immediately paid 99 per cent. to Shurltrust.

52. The statutory question was whether Shurltrust had received income or capital. If it was income, the effect of various tax avoidance provisions concerning the transfer of assets abroad was that the payment would be C deemed to be income of the McGuckians. If it was capital, the McGuckians would not be liable for tax. The McGuckians said that if Shurltrust had simply received the dividend, it would of course have been income. But Shurltrust did not receive the dividend. It received a payment from Mallardchoice which was a capital payment for an assignment of its right to income.

D

53. The Inland Revenue's argument, relying upon the formulation in the *Furniss* case [1984] AC 474; 55 TC 324 was that the assignment should be disregarded. The Northern Ireland Court of Appeal said (not, if I may respectfully say so, without justification) that one could not simply "disregard" the assignment. The payment of the money by Mallardchoice to E Shurltrust was the consideration for the assignment and an integral part of that transaction. If the assignment had to be disregarded, one could not explain how Shurltrust had received any money at all.

54. It seems to me that the Crown caused unnecessary difficulties for itself in the *McGuckian* case by failing to notice that the question was F different from that in *Furniss* v. *Dawson* and therefore did not necessarily respond to precisely the same analysis. In the *Furniss* case the question was the identity of the disponer. In the *McGuckian* case it was the nature of the payment received by Shurltrust—capital or income? In the former case, it is reasonable to speak of the middle stage of a chain of disposals being "disregarded". In the latter case, it makes much less sense. The question was G not whether the assignment should be disregarded but whether, from a commercial point of view, it amounted to an exchange of income for capital. Such exchanges usually have a commercial reality: the purchase or sale of an annuity, for example, is an exchange of capital for an income stream, involving a transfer of risk. But the transaction in the *McGuckian* case was nothing more than an attempt to relabel a sum of money. The fact that the H assignment had no commercial purpose did not mean that it had to be disregarded. But it failed to perform the alchemy of transforming the receipt of a dividend from the company into the receipt of a capital sum from someone else. For the purpose of the fiscal concept at stake, namely the character of the receipt as income derived from the company, it made no difference.

I

55. My Lords, I think that it was for these reasons that their Lordships in the *McGuckian* case went back to Lord Wilberforce's analysis in the *Ramsay* case [1982] AC 300; 54 TC 101 and tried to identify the principle of construction in play. Lord Steyn said that the decision marked a shift away from literalism to a "broad purposive interpretation" and from "formalistic insistence on examining steps in a composite scheme separately" to "a more

realistic legal analysis": [1997] 1 WLR 991, at pages 999–1000. Lord Cooke of Thorndon suggested, at page 1005; 69 TC 1, at page 84H that it was: A

> "an application to taxing Acts of the general approach to statutory interpretation whereby, in determining the natural meaning of particular expressions in their context, weight is given to the purpose and spirit of the legislation." B

56. My Lords, these are valuable insights and I respectfully suggest that particular attention should be paid to the way Lord Cooke of Thorndon 69 TC 1, at page 85C dealt with the criteria stated by Lord Brightman in *Furniss* v. *Dawson*;

> C "Lord Brightman spoke of certain limitations (a pre-ordained series of transactions including steps with no commercial or business purpose apart from the avoidance of a liability to tax). The present case does fall within these limitations, but it may be as well to add that, if the ultimate question is always the true bearing of a particular taxing provision on a particular set of facts, the limitations cannot be universal. Always one D must go back to the discernible intent of the taxing Act. I suspect that the advisers of those bent on tax avoidance ... do not always pay sufficient heed to the theme in the speeches in the *Furniss* case ... to the effect that the journey's end may not yet have been found."

57. I would only add that it is not only tax avoiders who may not pay E sufficient heed to the necessity of concentrating on the application of the particular taxing provision to the particular facts. The Inland Revenue sometimes also fails to do so. The journey's end may be different because the journey itself is not the same.

The limits of Ramsay. F

58. The limitations of the *Ramsay* principle therefore arise out of the paramount necessity of giving effect to the statutory language. One cannot elide the first and fundamental step in the process of construction, namely to identify the concept to which the statute refers. I readily accept that many expressions used in tax legislation (and not only in tax legislation) can be construed as referring to commercial concepts and that the courts are today G readier to give them such a construction than they were before the *Ramsay* case. But that is not always the case. Taxing statutes often refer to purely legal concepts. They use expressions of which a commercial man, asked what they meant, would say "You had better ask a lawyer". For example, stamp duty is payable upon a "conveyance or transfer on sale": see Sch 13, para 1(1) to the Finance Act 1999. Although slightly expanded by a definition in H para 1(2), the statutory language defines the document subject to duty essentially by reference to external legal concepts such as "conveyance" and "sale". If a transaction falls within the legal description, it makes no difference that it has no business purpose. Having a business purpose is not part of the relevant concept. If the "disregarded" steps in *Furniss* v. *Dawson* [1984] AC 474; 55 TC 324 had involved the use of documents of a legal I description which attracted stamp duty, duty would have been payable.

59. Even if a statutory expression refers to a business or economic concept, one cannot disregard a transaction which comes within the statutory language, construed in the correct commercial sense, simply on the ground that it was entered into solely for tax reasons. Business concepts have their

A boundaries no less than legal ones. Thus in two of the cases considered in *Craven* v. *White* [1989] AC 398; 62 TC 1 the House was unanimously of the view that although there had been an initial disposal with no commercial purpose, except to lay the ground for an avoidance of tax if and when there should be a further disposal to a third party, the transactions were so separate in fact as well as in law as to make it impossible to treat them, even B in a commercial sense, as a single disposal to the third party. The lapse of time between the two transactions, the lack of contemplation of any specific later disposal at the time of the first transaction, were commercial realities. The division of opinion in the House over how the third transaction should be categorised did not detract from the agreement that it had to fall within the statutory language.

C

60. Likewise the use of business concepts like "income" and "capital" may give the taxpayer a choice of structuring a commercial transaction so as to come within one concept or the other. As Lord Greene M.R. said in a celebrated passage in *Inland Revenue Commissioners* v. *Wesleyan and General Assurance Society* (1946) 30 TC 11, at page 16; [1946] 2 All ER 749, at page D 751:

> "In dealing with Income Tax questions it frequently happens that there are two methods at least of achieving a particular financial result. If one of those methods is adopted, tax will be payable. If the other method is adopted, tax will not be payable. It is sufficient to refer to the E quite common case where property is sold for a lump sum payable by instalments. If a piece of property is sold for £1,000 and the purchase price is to be paid in ten instalments of £100 each, no tax is payable. If, on the other hand, the property is sold in consideration of an annuity of £100 a year for ten years, tax is payable. The net result from the financial point of view is precisely the same in each case, but one F method of achieving it attracts tax and the other method does not."

61. It follows that a transaction which, for the avoidance of tax, has been structured to produce, say, capital, and does produce capital in the ordinary commercial sense of that concept (unlike the payment in *Inland Revenue Commissioners* v. *McGuckian* [1997] 1 WLR 991; 69 TC 1) cannot be G "recharacterised" as producing income: see *Commissioners of Inland Revenue* v. *Wattie* [1999] 1 WLR 873.

Tax mitigation and tax avoidance.

62. My Lords, it has occasionally been said that the boundary of the H *Ramsay* principle can be defined by asking whether the taxpayer's actions constituted (acceptable) tax mitigation or (unacceptable) tax avoidance. In *Inland Revenue Commissioners* v. *Willoughby* [1997] 1 WLR 1071, at page 1079 Lord Nolan described the concept of tax avoidance as "elusive". In that case, the House had to grapple with what it meant, or at any rate what its "hallmark" was, because the statute expressly provided that certain provisions should not apply if the taxpayer could show that he had not acted I with "the purpose of avoiding liability to taxation". The same question arises on the interpretation of the anti-avoidance provisions to which Lord Cooke of Thorndon referred in *Inland Revenue Commissioners* v. *McGuckian* [1997] 1 WLR 991, at page 1005; 69 TC 1. But when the statutory provisions do not contain words like "avoidance" or "mitigation", I do not think that it helps to introduce them. The fact that steps taken for the avoidance of tax are acceptable or unacceptable is the conclusion at which one arrives by applying

the statutory language to the facts of the case. It is not a test for deciding whether it applies or not. If I may be allowed to repeat what I said in *Norglen Ltd.* v. *Reeds Rains Prudential Ltd.* [1999] 2 AC 1, at pages 13–14:

> "If the question is whether a given transaction is such as to attract a statutory benefit, such as a grant or assistance like legal aid, or a statutory burden, such as income tax, I do not think that it promotes clarity of thought to use terms like stratagem or device. The question is simply whether upon its true construction, the statute applies to the transaction. Tax avoidance schemes are perhaps the best example. They either work (*Inland Revenue Commissioners* v. *Duke of Westminster* [1936] AC 1) or they do not (*Furniss* v. *Dawson* [1984] AC 474). If they do not work, the reason, as my noble and learned friend, Lord Steyn, pointed out in *Inland Revenue Commissioners* v. *McGuckian* [1997] 1 WLR 991, 1000, is simply that upon the true construction of the statute, the transaction which was designed to avoid the charge to tax actually comes within it. It is not that the statute has a penumbral spirit which strikes down devices or stratagems designed to avoid its terms or exploit its loopholes."

The present case: the appeal to Carnwath J. [1997] STC 1103; 73 TC 1(1)

63. My Lords, after what I fear was a lengthy analysis of the *Ramsay* principle I return to the present appeal. Carnwath J., who allowed an appeal from the Special Commissioners, said that the case was very much like *Inland Revenue Commissioners* v. *Burmah Oil Co. Ltd.* 54 TC 200; [1982] STC 30. In that case, the transaction left *Burmah* no worse off than it had been before and merely purported to convert a bad debt into an allowable loss. Similarly in this case, said Carnwath J., the transaction made no difference to the scheme or WIL but merely purported to convert an unpaid interest debt into a payment which could be deducted. In so doing, he treated the passage in the speech of Lord Diplock which I have already quoted (and the similar passage in the speech of Lord Brightman in the *Furniss* case) as being of general application, irrespective of the nature of the concept to which the statute refers.

64. My Lords, I can see that one could read these passages in such broad terms. But I do not think that it would be consistent with treating *Ramsay* as a principle of construction. In my opinion, what the *Burmah* case decided was that the statutory concept of a loss accruing upon a disposal has a business meaning and that the "disposal" and "loss" suffered by *Burmah* did not fall within it. To apply this reasoning to the present case, it would be necessary to construe the concept of payment in s 338 as having some business meaning other than the simple discharge of a debt. Otherwise one is not giving effect to the statutory language.

The Court of Appeal [1998] STC 1131; 73 TC 1(2)

65. The Court of Appeal unanimously allowed the appeal. Peter Gibson L.J. said that the question of whether the interest in this case had been "paid" could not be solved by using the same technique as had been used to decide whether there had been a "loss" in the *Burmah* case:

> "In other areas where fiscal legislation has provided for payment as an act significant for tax purposes and payment, within the meaning which the law would ordinarily give to it, has occurred, the courts have

(1) Page 22? *ante*. (2) Page 39? *ante*.

A been reluctant to accept that the *Ramsay* principle required the fact of payment to be ignored." [1998] STC 1131, at page 1144(1)

66. He referred to *Cairns* v. *MacDiarmid* (1982) 56 TC 556, at page 576 and *Customs and Excise Commissioners* v. *Faith Construction Ltd.* [1990] 1 QB 905. I shall return to these cases later. Peter Gibson L.J. said there was B nothing in the Taxes Act to suggest that "paid" should be construed to mean anything more than that the interest obligation had been discharged, wherever the money had come from. Pill L.J. delivered a concurring judgment and Mummery L.J. agreed.

The concept of payment

C 67. My Lords, payment of a debt such as interest ordinarily means an act, such as the transfer of money, which discharges the debt. It is accepted that in this case the interest debt was indeed discharged. So why did this not count as payment for the purposes of the Act? One of the difficulties which I have with the argument for the Crown is that I find the alternative concept D of payment for which it contends completely elusive. It is easy to understand a commercial sense of a loss which treats as irrelevant the fact that one part of a composite transaction produced a loss which was never intended to be more than momentary and theoretical. But what is the commercial concept of payment of a debt which treats as irrelevant the fact that the debt has been discharged? Mr. McCall does not contend that payment must involve a E negative cash flow which is not compensated by a cash flow in the opposite direction. He accepts, for example, that many commercial refinancing operations discharge old debts and create new ones without any cash flow either way. Nor is there any apparent policy to be found in s 338 which would require a negative cash flow. Otherwise, why should bank interest be deductible without any payment at all? As I have already said, the only F apparent reason for the insistence on payment of yearly interest is that payment gives rise to an obligation to deduct tax. In the present case, WIL complied with that obligation. The Crown's real complaint is that the scheme, as an exempt fund, was able to reclaim the tax. But this cannot be remedied by giving the word "paid" a different meaning in the case of a payment to an exempt lender. The word must mean the same, whatever the G status of the lender.

68. What the Crown finds objectionable is the circularity of the cash flow combined with the fact that the transaction took place entirely for tax purposes. And I accept that for the purposes of some concepts used in tax legislation, these two features would stamp the transaction as something H different from that contemplated by the legislature. For example, I have no doubt that Langley J. was right when he recently decided in *NMB Holdings Ltd.* v. *Secretary of State for Social Security* (unreported) 14 July 2000 that a payment of bonuses to directors in the form of platinum sponge held in a bank, accompanied by arrangements under which they could immediately sell it for cash to the bank, was not a "payment in kind" which fell to be I disregarded for the purpose of National Insurance Contribution. In commercial terms the directors were paid in money. It is obvious that such a transaction was not what the Social Security (Contributions) Regulations 1979 (SI 1979/591) contemplated as a payment in kind. But there can be equally little doubt that the bonuses were "paid" and, in the absence of some contrary context, I can see no reason not to treat them as paid when the

(1) Page 49D *ante*.

directors were credited with platinum sponge and the employer's obligation to pay them was discharged. A

The authorities

69. My Lords, like Peter Gibson L.J. in the Court of Appeal, I think that the taxpayer's case is supported by the authorities in which the concept of payment in a tax case has been considered. *Cairns* v. *MacDiarmid* 56 TC 556; [1982] STC 226 concerned an artificial scheme in which the taxpayer (Mr. Cairns) claimed to have paid £5,000 "annual interest" on a loan of £37,740 from his employer, Rossminster. The loan was in fact intended to last no more than four days. Mr. Cairns gave Rossminster a cheque for £5,000 in exchange for its cheque for £37,740. Nourse J. and the Court of Appeal held that the payment was not "annual interest" within the meaning of the Act. But the Crown also raised before Nourse J. the question of whether the £5,000 could be said to have been "paid". He said at 56 TC 556, at pages 576–577: B C

> "It is accepted by the Crown that the agreement between Mr. Cairns and Rossminster was not a sham, but a genuine transaction having the legal effect which it was expressed to have. On that footing I cannot see that Mr. Cairns ceased to 'pay' the £5,000 merely because he did so conditionally on receiving the £37,740 in exchange. If Mr. Potter's argument was correct it might have surprising results in its application to other genuine transactions. For example, there must be many loans where interest is payable in advance and where there is either, as there was in this case, an exchange of cheques or, perhaps more frequently, a payment by the lender to the borrower of a net amount representing principal less the first instalment of interest. It would be very strange if in either of those cases there was not a payment of interest for the purposes of [the Taxes Act], and to say that there was not would in my judgment attach to the word 'pays' a significance which in the context it cannot possibly bear." D E F

The other case to which Peter Gibson L.J. referred was *Customs and Excise Commissioners* v. *Faith Construction Ltd.* [1990] 1 QB 905. The question there was whether builders had received a "payment" in respect of a supply of services within the meaning of s 5(1) of the Value Added Tax Act 1983. That section provided that a supply of services was deemed to take place when the supplier received payment in respect of it. The facts were that in early 1984 a building company had entered into an agreement to erect a building but had not yet begun work. It was then announced in the March budget that with effect from 1 June 1984 the rate of VAT on building services would be increased from zero to the standard rate. To avoid payment of VAT, the customer paid the builder in advance. The builder then lent the money back to the customer on terms that it would be repayable only against architect's certificates for work done. The Commissioners of Customs and Excise, relying on the *Ramsay* case [1982] AC 300; 54 TC 101 argued that there had been no payment within the meaning of the Act or that if it had been, it was for the purposes of tax avoidance and should be "disregarded". The Court of Appeal said that there was no reason to construe "payment" in s 5(1) as meaning anything other than payment in discharge of the customer's obligation to pay for the services. Properly analysed, that obligation had been discharged and replaced by an obligation to repay money lent. Bingham L.J, said, [1990] 1 QB 905, at page 921: G H I

> "If we were entitled to disregard the legal effect of what was done here and give effect to the underlying substance, it might be possible to

A say that these payments were not really payments because they were made for the purpose of avoiding VAT and without any (or any other) commercial justification. But that is an approach which Lord Tomlin's well-known speech in *Inland Revenue Commissioners* v. *Duke of Westminster* [1936] AC 1, 19–21, roundly condemned where the transaction in question is genuine and I do not understand the principle B there laid down, described as 'cardinal' by Lord Wilberforce in *W.T. Ramsay Ltd.* v. *Inland Revenue Commissioners* [1982] AC 300, 323, to have been diluted or abrogated by later decisions. If the payments are to be disregarded the commissioners would, I think, have to show them to be a sham, and this they have not sought to do. If, as I have concluded, these were in law good contractual payments, then I do not think we are C entitled to disregard their legal effect and treat them as something else."

In other words, Bingham L.J. was saying that "payment" in s 5(1) was a legal concept and did not have some other commercial meaning. In my opinion the same is true of "paid" in s 338 of the Taxes Act.

D *Specific tax avoidance provisions*

70. The Revenue rely in the alternative upon three provisions which they say nullify the effect of the payment of interest. On all three I am in full agreement with the Court of Appeal and can therefore be very brief.

E (a) *Section 338(5)(a)*

71. This provides that a payment of interest under s 338(3) shall not be treated as a charge on income if it is "not ultimately borne by the company". There appears to be no case in which the meaning of this provision has been considered. It seems to contemplate some arrangement by which the burden of the interest payment is transferred to someone else. But there was no such F arrangement in this case. The burden of the interest payment never shifted from WIL. The Revenue submits that there was no burden because the interest payment was cancelled by the loan. This amounts to collapsing the two transactions and treating the interest as never having been paid at all. But this would be contrary to the findings of fact. Once it is accepted that the interest was paid, it seems to me that the burden of payment could only G have been borne by WIL.

(b) *Section 75(3)*

72. This provides that charges on income in a given accounting period can be carried forward to succeeding accounting periods only if they were paid "wholly and exclusively for purposes of the company's business." The H Revenue says that the interest payments were not paid for the purposes of the company's business but to make it more attractive to a purchaser. It was conceded that the loans upon which the interest was payable had been borrowed wholly and exclusively for the purposes of the company's business. The Special Commissioners said that one did not need to inquire into the purpose for which the taxpayer paid a legitimate debt which he had incurred for the purposes of his business. It is sufficient that the debt has been so incurred: see *Hyett* v. *Lennard* [1940] 2 KB 180. Like the Court of Appeal, I can see no error in this reasoning.

(c) *Section 787(1)*

73. This denies relief for payment of interest to a person who has paid pursuant to a scheme:

"... such that the sole or main benefit that might be expected to accrue to that person from the transaction under which the interest is paid was the obtaining of a reduction in tax liability by means of any such relief." A

74. The Revenue say that the interest was paid under a transaction from which the sole or main benefit which would accrue to WIL was the obtaining of a reduction in tax liability. Again, I have little to add to what the Court of Appeal said on this point. In my opinion it is plain that the "transaction under which the interest was paid" is the original loan and not the arrangements which enabled WIL to pay it. B

75. I would dismiss the appeal. For the reasons given by my noble and learned friend Lord Hope of Craighead, I would also dismiss the cross-appeal. C

Commissioner of Inland Revenue v Mitsubishi Motors New Zealand Ltd

10 Judicial Committee
10 July; 3 October 1995
Lord Keith of Kinkel, Lord Jauncey of Tullichettle, Lord Mustill, Lord Nicholls of Birkenhead and Lord Hoffmann.

15 *Revenue – Income tax – Deduction of expenditure under warranty against defects in motor vehicles – When expenditure incurred – Income Tax Act 1976, ss 101 and 104.*

The taxpayer sold motor vehicles through franchised dealers with the benefit of a
20 warranty against defects appearing within a year of delivery or until the vehicle had been driven for 20,000 km, whichever period was the shorter. The taxpayer agreed to indemnify the dealers against the cost of warranty claims. Under the terms of the warranty the vendor of the vehicle agreed to remedy defects if the defect appeared during the warranty period and if the purchaser notified the vendor
25 of the defect within 21 days of becoming aware of the defect.

The proper accounting treatment of the outstanding warranty liabilities was that as they were part of the cost of the vehicle sales, they should, so far as capable of reasonable estimation, be matched against the corresponding revenue. All vehicles which left the taxpayer's assembly plant were tested and examined for
30 defects. So far as the taxpayer was aware, there was nothing wrong with the vehicles but experience showed that in many cases a defect would be discovered during the warranty period. In the 1988 year 63 per cent of the vehicles sold were returned to the dealers for some kind of work to be done under a warranty. Although it could not be predicted whether any particular vehicle would turn out to be defective or
35 how serious the defect would be, the taxpayer could make a reasonably accurate forecast, based on previous experience, of what would be the total cost of remedial work for all the vehicles sold in a given year. Normal commercial practice therefore required that that amount should be brought into account as a deduction from income in estimating the profits or gains of the business in the year in which the
40 vehicles were sold.

Held: 1 The form in which the warranty had been expressed was not the end of the matter. Although the jurisprudential approach prevented one from treating an aggregate of contingent liabilities as a statistical certainty, it did not rule out
45 statistical estimation of facts which had happened but were unknown. The relevance of this principle was that estimation on the basis of statistical experience could be used to conclude that 63 per cent or thereabouts of the vehicles sold by the taxpayer in fact had defects which would manifest themselves within the warranty period. In deciding whether the taxpayer had incurred a liability at the time when the
50 vehicle was sold, it was legitimate to have regard to the evidence establishing that 63 per cent would in fact have defects (see p 518 line 31, p 518 line 55).

RACV Insurance Pty Ltd v Commissioner of Taxation [1975] VR 1; (1974) ALR 600 and *Commercial Union Assurance Co of Australia Ltd v Federal Commissioner of Taxation* (1977) 14 ALR 651; 7 ATR 435 approved.

Nilsen Development Laboratories Pty Ltd v Federal Commissioner of Taxation

of the Commonwealth of Australia (1981) 144 CLR 616; 33 ALR 161 and *Coles Myer Finance Ltd v Federal Commissioner of Taxation of the Commonwealth of Australia* (1993) 176 CLR 640 explained.

2 The language in which the warranty had been expressed made liability dependent upon the manifestation and notification of the defect within the 12-month period. The question of whether the taxpayer had been "definitively committed" to an expenditure or whether it was merely "impending, threatened or expected" did not depend upon whether future events which might determine liability were expressed in the language of contingency or defeasance. The question was rather whether, in the light of all the surrounding circumstances, a legal obligation to make a payment in the future could be said to have accrued. For this purpose, merely theoretical contingencies could be disregarded. That 63 per cent of the vehicles sold by the taxpayer had defects of a kind likely to manifest themselves within the warranty period was a matter of existing fact, not future contingency. Since these defects were by definition likely to show themselves within the warranty period, the contingency that the owners might be content not to require remedial work would be real only in the case of the most trivial defects. It would not make any material difference to the accuracy of the estimated amount of expenditure to which the taxpayer could be said, as a matter of law, to be definitively committed (see p 519 line 18, p 519 line 40).

Federal Commissioner of Taxation v James Flood Pty Ltd (1953) 88 CLR 492; 27 ALJ 481, *Coles Myer Finance Ltd v Federal Commissioner of Taxation of the Commonwealth of Australia* (1993) 176 CLR 640 and *Commercial Union Assurance Co of Australia Ltd v Federal Commissioner of Taxation* (1977) 14 ALR 651; 7 ATR 435 referred to.

Appeal dismissed.

Observation: The question of what income can be treated as "derived" during an accounting year is, unlike the question of deductions, a matter governed by normal accounting principles. If the taxpayer had actually made a separate charge for the warranty, there would be no difficulty about treating that income as earned over the warranty period rather than at the moment of sale. But there was no justification in the accounting evidence for retrospectively treating part of the sum agreed by the parties to be the price of the car as if it had been a separate charge for the warranty (see p 519 line 53).

Other cases mentioned in judgment

Bisley (A M) Ltd v Commissioner of Inland Revenue (1985) 8 TRNZ 513.

Inland Revenue (Commissioner of) v Farmers' Trading Co Ltd [1982] 1 NZLR 449 (CA).

Inland Revenue (Commissioner of) v Glen Eden Metal Spinners Ltd (1990) 12 NZTC 7,270.

Southern Railway of Peru Ltd v Owen (Inspector of Taxes) [1957] AC 334; [1956] 2 All ER 728.

Taxes (South Australia) (Commissioner of) v Executor Trustee and Agency Co of South Australia Ltd (1938) 63 CLR 108.

Appeal

This was an appeal against a decision of the Court of Appeal reported at [1994] 2 NZLR 392 on a case stated allowing an appeal of the taxpayer on an objection under the Income Tax Act 1976.

A Park QC (of the English Bar) and *P J H Jenkin QC* for the appellant.
A D MacKenzie and *R J Cullen* for the respondent.

Cur adv vult

The judgment of Their Lordships was delivered by

LORD HOFFMANN. Mitsubishi Motors New Zealand Ltd (MMNZ) sells motor vehicles through franchised dealers with the benefit of a warranty against defects appearing within a year of delivery or until the vehicle has been driven for 5 20,000 km, whichever period is the shorter. The question in this appeal is whether, in computing its "profits or gains" for the purpose of income tax, it can bring into account its anticipated liabilities under warranties remaining unexpired at the end of the year of account in which the vehicles were sold.

Income tax is levied upon "assessable income", which by s 65(2)(a) of the 10 Income Tax Act 1976 includes "all profits or gains derived from any business". The term "profits or gains" is not defined. Prima facie, therefore, it bears its ordinary meaning as it would be understood by a businessman or accountant. As Dixon J said in *Commissioner of Taxes (South Australia) v Executor Trustee and Agency Co of South Australia Ltd* (1938) 63 CLR 108, 152:

15

> "Income, profits and gains are conceptions of the world of affairs and particularly of business . . . in nearly every department of enterprise and employment the course of affairs and the practice of business have developed methods of estimating or computing in terms of money the result over an 20 interval of time produced by the operations of business, by the work of the individual, or by the use of capital. The practice of these methods of computation and the general recognition of the principles upon which they proceed are responsible in a great measure for the conceptions of income, profit and gain and, therefore, may be said to enter into the determination or 25 definition of the subject which the legislature has undertaken to tax."

The evidence of accounting practice adduced before Doogue J left no doubt about the proper treatment of the outstanding warranty liabilities. They were part of the cost of the vehicle sales and therefore, so far as capable of reasonable estimation, should be matched against the corresponding revenue. The evidence 30 satisfied the Judge that a reasonable estimate could be placed upon the anticipated liabilities. All vehicles which leave MMNZ's assembly plant at Porirua have been tested and examined for defects. So far as MMNZ is aware, there is nothing wrong with them. Nevertheless, experience shows that in many cases, a defect will be discovered during the warranty period. Often it is no more than a blemish in the 35 paintwork. Sometimes it is more serious. Sixty three per cent of the vehicles sold by MMNZ in the year 1988 were returned to the dealers for some kind of work to be done under the warranty. Although it cannot of course be predicted whether any particular vehicle will turn out to be defective or how serious the defect will be, MMNZ can make a reasonably accurate forecast, based on previous experience, 40 of what will be the total cost of remedial work for all the vehicles sold in a given year. Normal commercial practice therefore requires that this amount should be brought into account as a deduction from income in estimating the profits or gains of the business in the year in which the vehicles were sold.

The term "profits or gains" in s 65(2)(a) must however be read subject to the 45 provisions of ss 101 and 104 of the Income Tax Act 1976:

> **101. No deductions unless expressly provided** – Except as expressly provided in this Act, no deduction shall be made in respect of any expenditure or loss of any kind for the purpose of calculating the assessable income or the 50 non-assessable income of any taxpayer.
>
> **104. Expenditure or loss incurred in production of assessable income** – In calculating the assessable income of any taxpayer, any expenditure or loss to the extent to which it –
>
> (a) Is incurred in gaining or producing the assessable income for any income year; or

(b) Is necessarily incurred in carrying on a business for the purpose of gaining or producing the assessable income for any income year –

may, except as otherwise provided in this Act, be deducted from the total income derived by the taxpayer in the income year in which the expenditure or loss is incurred. 5

These sections have been construed to mean that in the calculation of the profits or gains of a business, the income side of the account is computed according to normal accounting principles but that "expenditure or loss" may be deducted only if it can be brought within the terms of s 104. In this respect the New Zealand 10 Income Tax Act 1976 differs from the United Kingdom Taxes Acts, which contain the equivalent of s 101 (s 817(1)(a) of the Income and Corporation Taxes Act 1988) but no equivalent of s 104. The United Kingdom Courts, faced with a statute which says that no deductions are to be allowed except those expressly enumerated and then fails to enumerate any permitted deductions, have felt able to treat the 15 concept of "profits or gains" as containing within itself a direction to make such deductions as normal accounting practice would require for the purpose of computing profits or gains. But s 104, which does expressly define the scope of the permitted deductions, makes it difficult to apply an accounting practice which is not in accordance with its express terms. In addition, although not for present 20 purposes relevant, there is s 106, which prohibits the deduction of various enumerated items of expenditure even if they come within s 104. It is not suggested that the warranty costs fall within any of the prohibited items and nothing more need therefore be said about s 106.

In construing s 104, the New Zealand Courts have followed Australian 25 authorities on the meaning of "losses and outgoings . . . incurred" in s 51(1) of the Income Tax Assessment Act 1936 and the equivalent provisions in earlier legislation. The phrase has been held to mean that the taxpayer must have either paid or become "definitively committed" to the expenditure. In *A M Bisley Ltd v Commissioner of Inland Revenue* (1985) 8 TRNZ 513, 528 Henry J summed up 30 the effect of the authorities, both Australian and New Zealand, in four propositions:

> "First, a particular expenditure is 'incurred' for tax purposes in an income year if it constitutes an existing obligation which arose in the course of that year. Second, where the expenditure arises under a written deed or agreement, whether or not it constitutes an existing obligation is a question of construction 35 of that deed or agreement. Third, that the expenditure is not payable until some future date does not of itself destroy its nature as an existing obligation. Fourth, that the expenditure is a defeasible liability does not of itself destroy its nature as an existing obligation."

40

There are two points about this construction which must be noted. First, it treats s 104 as concerned with *particular* items of expenditure rather than the aggregate sums which would concern a businessman drawing up his accounts. Each item, to be deductible, must satisfy the test of being an "existing obligation". This is to be contrasted with the normal commercial principles applied to the 45 computation of profits or gains. As Lord Radcliffe said in *Southern Railway of Peru Ltd v Owen (Inspector of Taxes)* [1957] AC 334, 357:

> "The answer to the question what can or cannot be admitted into the annual account is not provided by any exact analysis of the legal form of the relevant obligation . . . you get into a world of unreality if you try to solve your problem 50 in that way, because, where you are dealing with a number of similar obligations that arise from trading, although it may be true to say of each separate one that it may never mature, it is the sum of the obligations that matters to the trader, and experience may show that, while each remains uncertain, the aggregate can be fixed with some precision."

The second point is that the question of whether the expenditure has been "incurred" involves characterising the nature of the legal relationship between the taxpayer and the person to whom the obligation is owed. On one view, it requires one to decide as a matter of construction whether the obligation is contingent or
5 vested but defeasible. This is a nice distinction which can easily become a matter of language rather than substance and on which judicial views may differ; for an example, see *Commissioner of Inland Revenue v Glen Eden Metal Spinners Ltd* (1990) 12 NZTC 7,270. Both points illustrate the fact that this construction involves taking what the Australian Courts have called a jurisprudential rather than a
10 commercial view of the meaning of "incurred". This is an unusual approach to a taxing statute and Their Lordships detect in the Australian cases some degree of tension between loyalty to formal legal doctrine and reluctance to accept a computation of taxable profits which is wholly divorced from commercial reality.

Although there are clearly parallels between the Australian and New Zealand
15 legislation, there is also a striking difference. Unlike the New Zealand Act, the Australian Income Tax Assessment Act 1936 does not purport to levy tax upon "profits or gains". It defines "taxable income" in s 6 as "the amount remaining after deducting from the assessable income all allowable deductions". Section 25(1) defines "assessable income" as gross income and s 51(1) governs allowable
20 deductions:

> **51.** – (1) All losses and outgoings to the extent to which they are incurred in gaining or producing the assessable income, or are necessarily incurred in carrying on a business for the purpose of gaining or producing such income,
> 25 shall be allowable deductions except to the extent to which they are losses or outgoings of capital, or of a capital, private or domestic nature, or are incurred in relation to the gaining or production of exempt income.

Thus the Australian Act taxes the gross income less allowable deductions. Although both of these concepts necessarily involve some reference to accounting
30 principles, the statutory scheme does not apply the same principles to the relationship between them. In New Zealand, however, the use of the words "profits or gains" to describe the subject-matter of the charge to tax does prima facie require that income and expenditure should be related to each other in accordance with normal accounting principles. In *Commissioner of Inland Revenue v Farmers'*
35 *Trading Co Ltd* [1982] 1 NZLR 449, 456, Richardson J remarked upon this feature of the New Zealand legislative scheme, in relation to the Land and Income Tax Act 1954, as producing a "curious inconsistency". Mr MacKenzie, who appeared for the taxpayer, submitted that the Board should remove the inconsistency by departing from the Australian authorities and construing "expenditure or loss . . .
40 incurred" in s 104 to include all items of expenditure or loss which would be deductible on normal accounting principles. This was a submission which, in view of the weight of contrary authority in New Zealand, he had felt unable to make to the Court of Appeal. Their Lordships think that it is not without a certain attraction, but would be reluctant to adopt so revisionist an approach without a more thorough
45 inquiry into its possible repercussions on other parts of the legislation and the commercial practices which may have been designed around the established view of the law. Since Their Lordships feel able to dispose of this appeal on the basis of existing authority, they would prefer to keep the more fundamental point open.

Since the question of whether the warranty costs have been incurred within
50 the year in which the vehicle was sold is primarily a matter of construction, Their Lordships must set out the terms of the warranty. Strictly speaking, the warranty is given by the franchised dealer to the retail purchaser, but since MMNZ agrees to indemnify the dealers against the cost of warranty claims, any obligation incurred by the dealer will result in a simultaneous obligation being incurred by MMNZ. The warranty is as follows:

"NEW VEHICLE WARRANTY

1. THE vendor of the new vehicle described herein warrants to the original purchaser and subsequent owners that if in normal use and service during the relevant warranty period as provided below any defect appears in the material or workmanship of any part of the vehicle not otherwise warranted, and as soon as reasonably possible within 21 days of becoming aware of the defect, the purchaser returns the vehicle to the vendor's premises and notifies the vendor of the defect, the vendor will at the vendor's cost either (a) supply and fit, or (b) repair any such part acknowledged by the vendor to be defective. 5

2 THIS warranty shall not apply if the vehicle has been repaired or altered in any way other than by the vendor or in any service workshop not authorised by the vendor, or if the vehicle has been subjected to misuse neglect or accident, or if it has been loaded beyond manufacturer's loading capacity or operated in such a way that is not recommended by the manufacturer. 10

3. THE vendor shall not be liable for any loss or any consequential loss damage or expenses arising directly or indirectly from the defect. 15

4. THIS warranty is in lieu of all warranties terms conditions or representations expressed or implied whether by common law or statute.

5. THE new vehicle warranty period shall be 12 calendar months after delivery of the vehicle to the original purchaser or until the vehicle shall have run 20,000 km whichever first occurs." 20

There was a difference of opinion between Doogue J and the Court of Appeal over the effect of this warranty. Doogue J construed it as a promise that the vehicle was free from defects at the time of sale, so that in the case of a defective vehicle, liability was incurred at the moment of delivery. The purchaser's failure to comply with the conditions of the warranty was merely a ground upon which MMNZ might be able to avoid an existing liability. In the Court of Appeal, however, Richardson J held that any liability was contingent upon a defect appearing and being notified within the warranty period. Until then, no liability had been incurred. 25 30

On this point Their Lordships agree with the Court of Appeal. But they do not agree that the form in which the warranty is expressed is the end of the matter. There are two other principles which must also be taken into account. The first is that although the jurisprudential approach prevents one from treating an aggregate of contingent liabilities as a statistical certainty, it does not rule out statistical estimation of facts which have happened but are unknown. Thus in *RACV Insurance Pty Ltd v Commissioner of Taxation* [1975] VR 1 an insurance company carrying on accident business was allowed to make a deduction from its premium income of an estimated sum to represent its liabilities "incurred but not reported". These liabilities were not in law contingent. The accidents which gave rise to the company's liability had happened but the company did not know about them. A similar decision was reached in *Commercial Union Assurance Co of Australia Ltd v Federal Commissioner of Taxation* (1977) 14 ALR 651. Both cases were cited with approval in the High Court of Australia by Mason J (with whom Aickin J and Wilson J agreed) in *Nilsen Development Laboratories Pty Ltd v Federal Commissioner of Taxation of the Commonwealth of Australia* (1981) 144 CLR 616, 632. The learned Judge distinguished them from the cases on contingent liabilities because the accidents which gave rise to the liabilities under the policies had *occurred* during the relevant year of account. In the later case of *Coles Myer Finance Ltd v Federal Commissioner of Taxation of the Commonwealth of Australia* (1993) 176 CLR 640, 679, McHugh J remarked that the insurance cases involved a strained application of the earlier Australian decisions. This is true only in the sense that from a practical point of view, the distinction which they draw is irrelevant. But jurisprudentially the difference is clear enough. 35 40 45 50

The relevance of this principle is that estimation on the basis of statistical experience can be used to conclude that 63 per cent or thereabouts of the vehicles

sold by MMNZ in fact had defects which would manifest themselves within the warranty period of 12 months or 20,000 km. The finding of Doogue J on the evidence was that "63 per cent or thereabouts of all vehicles sold by [MMNZ] contain defects". Since this information could only be derived from MMNZ's experience of warranty claims, Their Lordships understand the finding to mean that this was the level of defects notified to dealers in accordance with the terms of the warranty. It also seems a fair inference that the defects were present at the time of sale. Mr Andrew Park QC, who appeared for the Commissioner, said that the terms of the warranty did not require that the defect should have existed at the time of sale. It could have come into existence within the warranty period. As a matter of construction, this is true. It is however hard to imagine the circumstances in which a defect in the "material or workmanship" of the vehicle would appear within 12 months of sale unless it was present, even if hidden, at the time the vehicle left the assembly plant. Any any rate Mr Park could not think of an example. In deciding whether MMNZ had incurred a liability at the time when the vehicle was sold, it is therefore legitimate to have regard to the evidence establishing that 63 per cent would in fact have had defects.

This, however, is not in itself enough to show that a liability was incurred. As has been said, Their Lordships agree with the Court of Appeal that the language in which the warranty was expressed made liability dependent upon the manifestation and notification of the defect within the 12-month period. But the Australian authorities show that the question of whether the taxpayer is "definitively committed" to an expenditure or whether it is merely "impending, threatened or expected" (to adopt the language used in the leading case of *Federal Commissioner of Taxation v James Flood Pty Ltd* (1953) 88 CLR 492, 506-507) does not depend simply upon whether future events which may determine liability are expressed in the language of contingency or defeasance. Their Lordships think it would be strange if a concept so eminently practical as the computation of profits for income tax depended upon theoretical distinctions more appropriate to the rule against perpetuities. The question is rather whether, in the light of all the surrounding circumstances, a legal obligation to make a payment in the future can be said to have accrued. For this purpose, merely theoretical contingencies can be disregarded. In *Coles Myer Finance Ltd v Federal Commissioner of Taxation* at pp 671-672, Deane J gave some examples of linguistic contingencies which were so unlikely as not to affect the certainty of the obligation. And in *Commercial Union Assurance Co of Australia Ltd v Federal Commissioner of Taxation* at pp 659-660, Newton J felt able to disregard a condition in an insurance policy requiring notice of the occurrence of an insured event to be given within a stipulated time on the ground that, according to the evidence, the condition was hardly ever insisted upon.

If one asks whether in respect of each of the vehicles sold by MMNZ, the warranty conditions make its liability contingent in substance as well as in form, the answer must be Yes. A substantial number – 37 per cent – will have no defects at all. But, for the reasons given above, Their Lordships think it legitimate to narrow the focus to those vehicles which left the assembly plant with defects of a kind likely to manifest themselves within the warranty period of 12 months or 20,000 km. That 63 per cent of vehicles had such defects was a matter of existing fact, not future contingency. Since these defects were by definition likely to show themselves within the warranty period, Their Lordships consider that the contingency that the owners might be content not to require remedial work would be real only in the case of the most trivial defects. It would not make any material difference to the accuracy of the estimated amount of expenditure to which the taxpayer could be said, as a matter of law, to be definitively committed.

Their Lordships would therefore respectfully differ on this point from the Court of Appeal and agree with Doogue J that the warranty costs were deductible under s 104. This makes it unnecessary to express any concluded view upon the

alternative basis upon which the Court of Appeal found for the taxpayer, namely that MMNZ could retrospectively apportion the price of each vehicle between the hardware and the warranty and then treat the warranty income as earned over the warranty period. Their Lordships are bound to say, however, that they have some difficulty with this method of reaching what is undoubtedly a sensible answer. The 5 question of what income can be treated as "derived" during an accounting year is, unlike the question of deductions, a matter governed by normal accounting principles. If MMNZ had actually made a separate charge for the warranty, there would be no difficulty about treating that income as earned over the warranty period rather than at the moment of sale. But there was no justification in the 10 accounting evidence for retrospectively treating part of the sum agreed by the parties to be the price of the car as if it had been a separate charge for the warranty. On the contrary, the taxpayer's expert accountant said that the whole price of the vehicles should be recognised as income and a deduction made for the warranty costs. It was on this basis that the statutory accounts were prepared. The fact that, 15 as the Court of Appeal held, the warranty costs could not be deducted for the purposes of income tax cannot in Their Lordships' view alter the accounting principles which govern the recognition of income. If upon its true construction the Income Tax Act 1976 requires the "profits or gains" of a business to be computed without deducting part of the cost of the business, it cannot be right, tempting as it 20 may be, to compensate for the anomaly by manipulating the ordinary rules for the recognition of income. Fortunately Their Lordships do not think that such measures are needed in this case.

The taxpayer also advanced a further alternative argument based upon the complicated group of ss 64B to 64M which are headed "Accrual Treatment of 25 Income and Expenditure Relating to Financial Arrangements". In view of the conclusion they have reached on the question of deductions, it is unnecessary for Their Lordships to say anything about this argument and they prefer not to. Their Lordships will humbly advise Her Majesty that the appeal ought to be dismissed. The Commissioner must pay the respondent's costs before Their Lordships' Board. 30

Appeal dismissed.

Solicitors for the appellant: *Crown Law Office* (Wellington).
Solicitors for the respondent: *Rudd Watts & Stone* (Wellington).

35

Reported by: Gloria Yee, *Barrister*

Introduction to Autopoietic Law

GUNTHER TEUBNER
Bremen, Firenze

Is the practice of legal reasoning bound to end in "strange loops", "tangled hierarchies", and "reflexivity dilemmas" (Hofstadter, 1979: 692; 1985: 70)? Is the legal process nothing but a closed cycle of recurrent legal operations: "computation of computation of computation..." (von Foerster, 1981: 296)? And are the social dynamics of the legal system based upon the "paradoxes of self-reference" (Wormell, 1958; Quine, 1976)? Up to now, the intricate problems of self-referential relations have not been part of the discourse of lawyers; they have been discussed outside the law, in logic, linguistics, cybernetics and general systems theory. Now the theory of legal autopoiesis is importing the logic of self-referentiality into the legal world. Legal autopoiesis breaks a taboo in legal thinking – the taboo of circularity. Legal doctrine, legal theory and legal sociology have all regarded circularity as a subject not to be broached. Circular arguments have been viewed as *petitio principii* forbidden by the iron law of legal logic. Legal autopoiesis now presumes to invalidate this iron law by transferring circularity from the world of ideas to that of hard facts. The message is that circularity is not a flaw in legal thinking which ought to be avoided (Fletcher, 1985: 1263), but rather that the reality of law consists of a multitude of circular processes.

Autopoiesis proposes, as a new and promising research strategy, to identify circular relationships within the legal system and to analyze their internal dynamics and their external interactions. What has been done so far in this field seems to be marginal. Legal methods are only slightly informed about the circular relation between purpose and norm in teleological reasoning (Alexy, 1978: 289 ff.); legal hermeneutics have only begun to study the perplexities of the hermeneutic circle in "*Vorverständnis und Methodenwahl*" (Esser, 1970); legal sociology has so far granted the luxury of circularity in feedback loops only between law and society. In the autopoietic perspective, these phenomena are viewed as only a few special cases in the encompassing circular self-referential reality of law. The whole legal system is seen as a dynamic cyclical reproduction of legal elements embedded in hypercyclical relations of legal structures and processes (see

LAW LIBRARY
VICTORIA UNIVERSITY OF WELLINGTON

Teubner, 1987 a). Law, like other autopoietic systems, is nothing but an "endless dance of internal correlations in a closed network of interacting elements" (Maturana, 1982: 28).

Circularity suggests closure. Thus, an autopoietic legal system is seen as operationally closed. Apparently, this runs against modern conceptions of an open responsive legal order, adapting to, and at the same time, shaping the social environment. Operational closure of law raises the suspicion of a new legal formalism, of a self-sustaining autarchy of law, of a new ideology for the legal profession — especially in the minds of legal sociologists. However, operational closure is only a half-truth. "L'ouvert s'appuye sur le fermé" (Morin, 1977: 197 ff.). A radical closure of the system — under certain conditions — means its radical openness. This is one of the most challenging theses of autopoietic theory. The more the legal system gains in operational closure and autonomy, the more it gains in openness toward social facts, political demands, social science theories, and human needs. This sounds paradoxical, and many formulations within this new theory seem to enjoy paradoxes. However, if it can be shown that this combination of openness and closure renders the legal system more responsive to social reality, and if concrete mechanisms coupling normative closure with cognitive openness can be identified, then legal autopoiesis has a potential of going beyond the theory of "open systems" which views itself as the legitimate successor of the older concepts of closed systems.

This book launches an inquiry into the world of legal autopoiesis — it is a project in interdisciplinary theory building. Our collective effort is to transfer a new and complex theory, the theory of autopoietic systems, from the general context of systems theory to the context of legal theory. This is done with the cooperation of international experts in the fields of both legal theory and systems theory. The book represents the first major endeavor to develop this approach within the field of law.

What sense is there for legal theory to make use of the theory of autopoietic systems? This newly developed theory, formulated by biologists (Maturana, 1970; Maturana et al., 1974; Varela, 1979; Zeleny, 1981; Roth and Schwegler, 1981) and transferred to the social sciences (Hejl, 1982 a; 1982 b; Luhmann, 1981; 1984; 1985; Teubner, 1985; 1987 a; 1987 b; Teubner and Willke, 1984) cannot yet claim with authority to be a fruitful paradigm. It is thus used in a more experimental manner as a strictly heuristic device. What follows for the problematic law and society relation if it is reformulated in terms of autopoiesis and self-referentiality? What hypotheses and consequences are implied with regard to doctrinal practice and legal policy?

While the relations between law and other spheres of society will be dealt with in a separate volume (Teubner, 1987 c), the *idée directrice* of this book is to examine the potential of autopoiesis for legal theory and to reformulate fundamental legal concepts in the light of this theory. Although

it is tempting to start with a critical assessment of autopoiesis as such, it seems to be a more fruitful approach if we first explore the changes in legal theory suggested by the concept, and then follow up with a critical examination of the concept and its legal consequences. Thus, the book is organized as a debate between protagonists and critics of autopoiesis, concentrating on the following issues:

- Legal System
- Legal Autonomy
- Legal Change
- Legal Cognition

1. Legal System

What is autopoiesis? An autopoietic system produces and reproduces its own elements by the interaction of its elements. According to *Niklas Luhmann*, the main representative of social autopoiesis, the decisive innovation in comparison to older theories of self-organization is that certain systems are capable not only of creating an autonomous order but of creating their own elements as well. Autopoietic systems are not only identified in biology as cells and organisms, on the basis of life, but they can also be identified as social systems (interaction, organization, society), on the basis of meaning. The basic element of a social system is communication – not the human being. Communication as the unity of utterance, information and understanding constitutes social systems by recursively reproducing communication. This conceptualization of social systems brings about three important changes in social theory: (1) a radical temporalization of social systems – this means that the elements are not "stable" but are "events" which necessitate the system to recursively use events to produce events; (2) a reformulation of the old maintenance problem: what has to be maintained is the recursively closed organization of an open system; closure as a condition for openness is one of the central tenets of autopoiesis; (3) epistemological consequences: for the theory of autopoiesis, observation is an activity which can be performed by living systems, psychic systems and social systems. Observation is itself an operation of an autopoietic system.

If society is an autopoietic system of recursively self-reproducing communication, can one then also identify the law as being an autopoietic system of its own? In other words: Is there a legal autopoiesis, beyond social autopoiesis? According to Luhmann, in his essay, "The Unity of the Legal System", the basic units of a legal system are neither, as lawyers are used to describing them, legal norms, nor, as sociologists define them, actors and organizations – law is a system of communications. However, legal

autopoiesis is brought about only if an emergent element of the legal system is created: the legal act. Legal acts are those communicative events that change legal structures. Here we find the basic circularity that defines the legal system: the circular relationship between legal acts and legal norms. Circularity replaces extra-legal foundations of law as does Kelsen's fictitious *Grundnorm*. If the recursive reproduction of legal acts constitutes legal autopoiesis, then operational closure is the main condition for its environmental openness. This interplay of openness and closure is represented in the legal system by the combination of normative closure and cognitive openness. The important consequence is that only a limited scope exists for the combination of normative expectations with cognitive expectations. This, according to Luhmann, is the key to the explanation of the modern crisis phenomena of law. Finally, autopoiesis of law limits, to a considerable degree, the variability of legal functions. The autopoietic closure sets effective limits to the political instrumentalization of law.

What is the significance of legal autopoiesis in the history of legal thought and what is its role in modern society? "The Law of Law" is *François Ewald*'s answer to this question by which he attempts a contextual interpretation of legal autopoiesis. He deals with two issues: What role does autopoiesis play in the conflict between legal theory and legal sociology? Does autopoiesis supply an adequate theory for the law of the modern age which Ewald calls the "social law"?

In the contemporary dilemmatic schism between formal analytical theories of law and more empirically-based legal sociology, legal autopoiesis appears in Ewald's view as a kind of "Columbus' egg". Legal autopoiesis is at the same time post-Kelsen and post-sociological. Both the internal viewpoint of the Kelsenian tradition which has been unable to account for differences between legal systems, and the critical project that inspired the external viewpoint of legal sociology are in deep crisis today. In this historical situation of legal thought, autopoiesis is emerging as an ambitious attempt to reconcile the antagonistic views. Autopoiesis transcends and conserves at the same time the split between pure theory and sociology. The view of law as a self-reproductive system of communication, the legal act as the element of the legal system, the combination of normative closure and cognitive openness – these features of legal autopoiesis give it a chance, Ewald maintains, to overcome the contemporary crisis in legal theory and sociology.

In Ewald's view the significance of legal autopoiesis is not confined to its role in legal thought. What is at stake is an adequate theory of law under conditions of modernity which Ewald characterizes by the destruction of natural law concepts, the loss of universalism, the particularization of political sovereignty and the relativization of science. Under these circumstances law can no longer be founded on principles outside itself. Legal practice in the twentieth century, according to Ewald, has found a solution

to this need for self-foundation: the formulation of principles of law based exclusively on legal reflexivity. Legal autopoiesis may well serve as a theoretical re-formulation of those factual developments in modern legal practice which Ewald describes as "social law".

But is autopoiesis in a position to act as a critique of the legality of social law? Ewald argues that as long as autopoiesis seeks to develop in the traditional forms of theoretical discourse, it is condemned to reformulate empty tautologies. The practice of social law is in need of a self-description and self-critique that makes objectivity of legal judgment possible but avoids at the same time universal statements: objectivity without universality. Ewald asserts that legal autopoiesis must take on a "resolutely atheoretical form" if it is to become the critical legal philosophy of our time.

Jean-Pierre Dupuy continues these efforts to locate legal autopoiesis in the context of contemporary social and legal philosophy by confronting it with the normative theories of Robert Nozick and Friedrich von Hayek. He takes up the paradox of closure and openness in autopoietic systems and supplies three different interpretations of how the legal system can be open and closed at the same time: (a) closure/openness refer to different domains of the legal system (normatively closed, but cognitively open); (b) legal closure implies legal openness ("order from noise"); and (c) self-transcendence of a normative order, the distancing of law in relation to itself, what he considers to be the most subtle form of openness. His main thesis is that a unit is self-transcending insofar as the complexity of the behavior of which it is capable is infinitely superior to the control capacities of the unit. In this perspective, Dupuy analyzes Nozick's theory of entitlement with its self-referential character of pure procedural justice and Hayek's theory of law as a "spontaneous" social order. In both theories, he identifies important parallels to the constructs of self-reference and autopoiesis. Dupuy ends with a paradoxical interpretation of legal autopoiesis locating it at the "height of modernity": If, at the same time, one sees the irreducible *Grundlosigkeit* of law and its very necessity, the "squaring of the circle" is to posit that law is self-legitimizing and self-founding. "The autonomy of the legal system is, finally, undecidably, decided".

The dialectics of legal openness and closure are taken up by *François Ost* from a different angle — from the angle of the individual actor in the perspective of game theory. He criticizes the autopoietic paradigm in its almost autistic closure by contrasting it with three positions. (1) Ashby's theorem: A program programming itself is a logical impossibility; changes must be produced from the outside; (2) Atlan's *auto-organisation relative*: according to the "order from noise" principle the internally undetermined system transforms external perturbations into an organizational order; (3) Castoriadis' *société autonome*: a society is autonomous when it is constantly able to free itself of the laws, values and traditions which it has imposed on itself. Ost develops a model of *le jeu du droit* in which closure is

represented by regulation and convention while openness and dynamics are introduced into the system by individual actors and their strategies.

One of the crucial steps in legal autopoiesis is the transfer of biological models into social and legal spheres. *Hubert Rottleuthner* analyzes and criticizes the use of "Biological Metaphors in Legal Thought". He asks the question: How are concepts changed when they are transferred from one field to another and how does a biological metaphor allow the solution of major problems in social and legal contexts? He begins with an historical analysis of the conceptual transfer describing how the concepts of "organism" and "evolution" have been used in legal thinking. He then critically assesses the transplantation of autopoiesis into the field of law. He identifies three areas in which the concept of autopoiesis can help solve specific problems: (1) A new concept of law can be explicated which would enable hitherto disparate elements of legal theory to be integrated or seen in a different light. In this context, Rottleuthner examines the concept of autopoiesis by referring to the legal theories of Kant, Kelsen and Hart; (2) Legal regulation processes can be explained by pointing to the limits of law's capacity for social control and identifying information requirements for legal control. According to Rottleuthner, autopoiesis helps identify autonomous social structure and organization. What is missing, however, is a more concrete analysis of regulated systems; (3) Theories of evolution are reinterpreted in terms of social and legal change. Autopoiesis stresses the role of an internal evolution of law. Again, Rottleuthner insists on a more empirically-oriented formulation of the theory of autopoiesis, which would lead to the identification of concrete evolutionary mechanisms.

2. Legal Autonomy

Legal autopoiesis is probably most controversial in its insistence on legal autonomy. After all, modern law is shaped by external constraints, social pressures and political decisions. How can one, after the social science revolution in law, after sociological jurisprudence and legal economics, still describe the legal system as autonomous (Friedman, 1985)? *Hans-Georg Deggau*, in his essay on the "Communicative Autonomy of the Legal System", carefully elaborates the conditions of the possibility of legal autonomy based upon legal communications as elementary units of a legal system. Deggau identifies legal autonomy in the "conditionalized legal normativity" which is capable of producing the necessary surplus for autopoietic closure. Normativity, as the basis of legal autonomy, produces, at the same time, the specific relation of law to other social systems, since, in its normative structures, law has a "structural affinity" towards other social systems. The very function of law is to congruently generalize normative structures in its social environment.

This is a concept of legal autonomy which stands in sharp contrast to classical socio-legal assumptions about the interrelations between law and society. It is precisely this difference which *Richard Lempert* focuses on in his critique of legal autopoiesis. He constructs two models of legal autonomy. An "Anglo-American vision" as an empirically-oriented theory about actual operations of law is contrasted with a "Continental vision" which is abstract – developed by logic, analogy and a firm *a priori* theoretical commitment to a model. While the latter stems from a general concept of autopoietic legal autonomy as a radical operational closure, the Anglo-American vision is able to distinguish varying degrees of autonomy by defining empirically testable criteria of relative autonomy. From this clear-cut juxtaposition it becomes evident that autonomy of the law in the autopoietic sense is either "palpably wrong" or "trivial". This is a harsh statement. Perhaps it tells something about the relationship between "autopoiesis" and "law and society" as the well-known incompatibility between competing pradigms.

This is precisely the point where *David Nelken* picks up the argument. He interprets the controversy between Luhmann and Lempert on legal autonomy as a conflict indicative of a paradigm change. This enables him to point to radical differences in the conception of legal autonomy, in regard to its theoretical framework, its characteristic meaning, its policy concerns, and methods of investigation. His main point, however, is to show that legal autopoiesis represents a full-fledged paradigm change from mainstream law and society. For this purpose, Nelken draws an intellectual map of competing socio-legal approaches placing "law and society" at the intersection of two axes, one dealing with the relationship between law and society (conceived ontologically), the other with that between law and sociology (conceived as epistemologies). Nelken sees the great merit of autopoiesis theory to collapse these axes into each other by showing how the extremes do in fact meet up. Two consequences are important: autopoietic theory shows how, where and why law, society and sociology actually became distinguished from each other as an evolutionary outcome of functional differentiation. At the same time autopoiesis theory resists characterizing law as a mere passive object of the sociological gaze and treats it rather as a subject reproducing its own conceptions of itself and of society.

3. Legal Change

If the identity of an autopoietic system is defined by its organizational closure, by the recursive reproduction of its elements, then autopoiesis appears to be resistant to change, learning and evolution. However, if one takes into account the fundamental distinction between autopoietic

organization, which is kept constant, and concrete structures of the system, which can change within the boundaries of the autopoietic organization (Maturana, 1982: 282 f.), then it becomes clear that autopoiesis does not exclude evolution but implies a redefinition of evolution. *Gunther Teubner*, in his essay on "Evolution of Autopoietic Law", attempts to resolve two problems of autopoietic evolution. How does the legal system evolve into autopoietic closure? How does legal evolution operate after an autopoietic closure of the legal system? The first question of pre-autopoietic evolution is answered with reference to the construction of a "hypercycle" (Eigen and Schuster, 1979). It is by mechanisms of "blind" evolution that "socially diffuse law" develops higher forms of autonomy via the cyclical constitution of its system components. It gains autopoietic closure by the hypercyclical connection between them. The answer to the second question of post-autopoietic evolution is "internalization". Autopoietic law is separated from the general evolution of its social environment and develops internal legal mechanisms for the evolutionary functions of variation, selection and retention. At the same time, legal development is coupled to broader social developments by specific mechanisms of co-evolution.

Karl-Heinz Ladeur's contribution is also concerned with legal change, focusing on the adaptability of law under modern conditions. Ladeur asks if and to what degree the functions of modern law, understood as an autopoietic system, can vary. He criticizes Luhmann's version of legal autopoiesis as too static in regard to the variability of functions, and develops a three-level-model of the relationship between legal structures and functions. Social functions of law change through the interaction of those three levels: (I) the organized level (structure – subsumption model in the classical sense); (II) the level of the "organizing function" (controlling transformations on level I); (III) the level of "virtualizing" structure and function. On this level potential destabilization leads to "strange loops" intertwining all three levels with the result that fluctuations of level I, being mitigated on level II, are strengthened on level III and lead to a change of function of level II. More concretely, Ladeur observes not only structural but functional changes in modern law which point to an emerging function of law that could consist in securing an abundance of alternatives, openness, flexibility for a new legal rationality of bargaining and compatibilization as opposed to the classical function of law which consisted in congruently generalizing social expectations.

While Ladeur and Teubner stress the internal dynamics of legal evolution, *Thomas Heller* focuses on the co-evolution of different autopoietic systems, more precisely, on the evolutionary dynamics in the interaction between law, economy and culture. Drawing on the Hispano-American biculturalism controversy in the United States, Heller stresses the intersystemic origins of perturbations of the legal system and develops a model of autopoietic co-evolution, based on the assumptions of self-referential reproduction of

core practices, system closure, instability and locality. The goals of the analysis are: (1) to define precisely the reproductive autopoietic core in the legal system; (2) to describe the interactive effects among co-existent systems, law, economy and culture; (3) to explain the internal dynamics of self-reproduction that create, at the same time, variation under the influence of external perturbations. Heller construes a multitude of cyclically organized system-internal "matching processes". Under the influence of environmental perturbations, the legal system develops a heuristic model of processed information as legal claims and matches it against existing legal practices. In the negative case this leads to the development of "causal models" and even "causal models of causal models" which in turn are used for reworking the system's prefiguration rules. Heller discusses the political relevance of such a cyclical model as a problem of "the politics of autopoiesis". In opposition to both neo-liberal and structuralist theories, autopoiesis is situated in a post-structuralist discourse, emphasizing the complex, the local, the closed and the unstable against global, coherent, artificing and equilibrating mechanisms.

4. Legal Cognition

Legal autopoiesis suggests to rethink questions of legal epistemology, especially in its insistence on normative closure and cognitive openness. "The Fact of Law" is one of the most important epistemological questions that *Patrick Nerhot* takes up. He contrasts traditional realist positions in law, especially positions of legal positivism, with a "radically constructivist" position which is suggested by legal autopoiesis. Legal facts are not imported into the law from the outside, they do not represent an independent reality, but they are constructed within the law by operations of the legal system. This view does not render the distinction between law and fact meaningless, but it reinterprets and connects this distinction to the difference between normative and cognitive operations.

Niklas Luhmann, in his concluding essay "Closure and Openness: On Reality in the World of Law", takes up themes of legal epistemology in order to answer to the objections which have been raised during the proceedings of the conference. The main issues are: How can one dissolve the paradox that law is at the same time an open system and a closed system? How can one insist on legal autonomy and stress at the same time its inter-dependencies with society at large? How is co-evolution of law and society conceivable under the premise that law as well as society are autopoietic systems of their own? How can one account for the apparent contradiction that the legal system is constructing its own reality by

recursive legal operations and is at the same time exposed to the constraints of extra-legal reality?

Luhmann approaches these interrelated questions first by posing a dilemma and second by proposing a solution. The theory of self-referential systems, he maintains, leads inevitably to the following dilemma: on the one hand, no system is in a position to operate outside its boundaries; no system has a direct access to reality "out there". On the other hand, the rate of structural evolution requires the assumption that a system's environment is non-random and structured and produces effective constraints for the system. An epistemological theory must be in a position to resolve this dilemma. Luhmann's proposal consists in a distinction. He distinguishes between internal information processing and external constraints via a materiality continuum and via simultaneous presence of events.

Luhmann insists on the strictly system-specific and exclusively internal character of information processing. A transfer of information from the environment to the system and vice versa is impossible. Via information the environment is inaccessible to the system. However, the environment effects its presence upon the system by imposing drastic constraints which limit the system's operational potential. Luhmann identifies two mechanisms of environmental coupling. One mechanism he calls materiality continuum. It is the material-energetic "basis" of meaning systems which exerts considerable constraints on them. The second mechanism is the simultaneous presence of events in several meaning processing systems. This allows for the interpenetration of autonomously operating mechanisms.

With this dualism of information and interpenetration, Luhmann attempts to answer the above-mentioned questions. Legal operations cannot reach beyond the boundaries of law. In this respect law is a closed system. In its cognitive operations law still can operate only within its boundaries, but it opens itself to its environment by constructing a legal reality. Beyond those mechanisms of normative closure and cognitive openness, the law is coupled to its social environment via mechanisms of interpenetration. Law participates in language structures and reality constructions of general social communication. And it is connected to other spheres of society by the simultaneous presence of social events which are selectively processed in law and in its environmental subsystems.

This implies a new understanding of legal autonomy. Autonomy is no longer internal causal determination of the system's operations. Legal autonomy is the selective combination of legal operations with legal operations. This autonomy is compatible with political and economic pressures on the legal system. Autonomy is in danger only once the legal coding as such is in danger of being replaced by criteria of economic utility and political expediency.

Bibliography

ALEXY, ROBERT (1978) *Theorie der juristischen Argumentation. Die Theorie des rationalen Diskurses als Theorie der juristischen Begründung*. Frankfurt a. M.: Suhrkamp.

ESSER, JOSEF (1970) *Vorverständnis und Methodenwahl in der Rechtsfindung: Rationalitätsgarantien der richterlichen Entscheidungspraxis*. Frankfurt a. M.: Athenäum.

EIGEN, MANFRED and PETER SCHUSTER (1979) *The Hypercycle: A Principle of Natural Self-organization*. Berlin: Springer.

FLETCHER, GEORGE P. (1985) "Paradoxes in Legal Thought", 85 *Columbia Law Review* 1263.

FOERSTER, HEINZ VON (1981) *Observing Systems*. Seaside: Intersystems Publications.

FRIEDMAN, LAWRENCE (1985) *Total Justice*. New York: Russell Sage.

HEJL, PETER M. (1982 a) "Die Theorie autopoietischer Systeme. Perspektiven für die soziale Systemtheorie", 13 *Rechtstheorie* 45.

— (1982 b) *Sozialwissenschaft als Theorie selbstreferentieller Systeme*. Frankfurt a. M.: Campus.

HOFSTADTER, DOUGLAS R. (1979) *Gödel, Escher, Bach: An Eternal Golden Braid*. New York: Basic Books.

— (1985) "Nomic: A Self-Modifying Game Based on Reflexivity in Law" in D. Hofstadter, *Metamagical Themes*. New York: Bantam.

LUHMANN, NIKLAS (1981) "Evolution des Rechts" in N. Luhmann, *Ausdifferenzierung des Rechtssystems. Beiträge zur Rechtssoziologie und Rechtstheorie*. Frankfurt a. M.: Suhrkamp.

— (1984) *Soziale Systeme. Grundriß einer allgemeinen Theorie*. Frankfurt a. M.: Suhrkamp.

— (1985) "The Self-Production of Law and its Limits" in G. Teubner (ed.) *Dilemmas of Law in the Welfare State*. Berlin: de Gruyter.

MATURANA, HUMBERTO R. (1970) *Biology of Cognition*. Urbana.

— (1982) *Erkennen: Die Organisation und Verkörperung von Wirklichkeit*. Braunschweig: Vieweg.

MATURANA, HUMBERTO R., FRANCISCO E. VARELA and R. URIBE (1974) "Autopoiesis: The Organization of Living Systems, Its Characterisation and a Model", 5 *Bio Systems* 187.

MORIN, EDGAR (1977) *La méthode*, Vol. 1, *La nature de la nature*. Paris: Seuil.

QUINE, WILLARD V. (1976) *The Ways of Paradox*. Cambridge: Harvard University Press.

ROTH, GERHARD and HELMUT SCHWEGLER (eds.) (1981) *Self-organizing Systems. An Interdisciplinary Approach*. Frankfurt a. M.: Campus.

TEUBNER, GUNTHER (1985) "After Legal Instrumentalism? Strategic Models of Post-Regulatory Law" in G. Teubner (ed.), *Dilemmas of Law in the Welfare State*. Berlin: de Gruyter.

— (1987 a) "Hypercycle in Law and Organization", *European Yearbook for Legal Sociology*.

— (1987 b) "Social Order from Legislative Noise" in G. Teubner (ed.), *State, Law, Economy as Autopoietic Systems*. Berlin: de Gruyter.

— (1987 c) *State, Law, Economy as Autopoietic Systems*. Berlin: de Gruyter (forthcoming).

TEUBNER, GUNTHER and HELMUT WILLKE (1984) "Kontext und Autonomie. Gesellschaftliche Selbststeuerung durch reflexives Recht", 5 *Zeitschrift für Rechtssoziologie* 4.

VARELA, FRANCISCO J. (1979) *Principles of Biological Autonomy*. New York: Elsevier.

WORMELL, C. P. (1958) "On the Paradoxes of Self-Reference", 67 *Mind* 267.

ZELENY, MILAN (1981) *Autopoiesis: A Theory of Living Organization*. New York: North Holland.

The Theory of Autopoiesis as an Approach to a better Understanding of Postmodern Law

- From the Hierarchy of Norms to the Heterarchy of Changing Patterns of Legal Inter-relationships -

Professor Dr. Karl-Heinz Ladeur

EUI WORKING PAPERS

1. FOREWORD 6

2. THE CONCEPT OF „SYSTEM“ IN TRADITIONAL LEGAL APPROACHES AND IN SYSTEMS THEORY 7

3. WHAT IS AUTOPOIESIS? 9

A. "CLOSURE" OF THE SYSTEM - ANIMADVERSIONS ON THE CONCEPT OF 'AUTONOMY' 9
B. SYSTEMS AS 'HISTORICAL MACHINES' OPERATING WITH DISTINCTIONS 12
C. SYSTEMS ONLY EXIST IN THE PLURAL FORM 13

4. THE LEGAL SYSTEM - SOME BASIC THEORETICAL ELEMENTS 13

A. THE GUIDING DISTINCTION OF THE LEGAL SYSTEM 13
B. THE DISTINCTION OF CODE/PROGRAMME AND CENTRE/PERIPHERY IN THE LEGAL SYSTEM 14

5. FROM A UNIVERSALISTIC TO A RELATIONAL PARADIGM OF LAW 15

A. WHAT IS THE BASIS OF UNIVERSALISTIC LAW? 15
B. WHY RELATIONSHIPS MATTER 16

6. THE CHALLENGE OF LEGAL PLURALISM 18

A. STANDARDS, THE NEW LAW MERCHANT AND THE GLOBALIZATION OF LAW 18
B. IN SEARCH OF A „META-DOCTRINE“ OF LEGAL PLURALISM 21

7. THE RELATIONSHIP BETWEEN LIBERAL CONCEPTIONS OF LAW AND PRACTICAL KNOWLEDGE 23

A. ORIENTATION PROBLEMS OF THE LAW IN CLASSICAL LIBERALISM AND POST-MODERNISM 23
B. TAKING BOUNDED RATIONALITY SERIOUSLY? 26

8. HOW DOES SYSTEMS THEORY DEAL WITH THE TRANSFORMATION OF LAW IN POST-MODERNITY? 28

A. ARE COURTS REALLY THE CENTRE OF THE LEGAL SYSTEM? 28
B. THE ORIENTATION PROBLEM OF COURTS 29
C. AUTOPOIETIC THEORY AS A THEORY OF POST-MODERNITY 32
D. FROM THE HIERARCHICAL RELATIONSHIP BETWEEN LAW AND ITS

CONCRETE APPLICATION TO THE HETERARCHICAL COOPERATION OF LAW AND THE SELF-ORGANIZATION OF SOCIETAL LEGAL TRANSACTIONS IN A "NETWORK OF NETWORKS" 33

9. GLOBAL LAW BEYOND THE STATE 36

A. EVOLUTION OF TRANSNATIONAL FORMS OF LAW 36

B. LAW BEYOND THE STATE: THE EXAMPLES OF THE NEW LAW MERCHANT, THE SUPRANATIONAL CHARACTER OF EUROPEAN LAW AND THE RISE OF WTO LAW....38

C. FROM EXPERIENCE-BASED LAW TOWARDS COOPERATIVE LAW FOR THE MANAGEMENT OF UNCERTAINTY 42

10. CONCLUSIONS 44

The Theory of Autopoiesis as an Approach to a better Understanding of Postmodern Law

- From the Hierarchy of Norms to the Heterarchy of Changing Patterns of Legal Inter-relationships -

Professor Dr. Karl-Heinz Ladeur

1. Foreword

Trying to explain the autopoietic theory of law to readers who I suppose not to be familiar with its constructions can put one in an awkward position because some of the elements of autopoietic theory of law are based on the autopoietic theory of society. But this again needs some explanation which I will try to give. In addition, I will try to confirm my hypothesis in some respect to Robert Alexy's presentation of a discourse-ethical theory of law which might limit the confusion to an acceptable measure.

First, I will give an idea of the concept of systems based on rules and sentences that we, as lawyers, are familiar with. Second I shall present autopoietic theory, as such, in a nut shell - a hazel nut shell. Third, I shall give an outline of the theory of autopoietic law in a nut shell - a walnut shell. Fourth, I shall try to explain what the rationale of this theory, as opposed to the competing theories of Habermas, Alexy etc., really is. And finally, I shall try to explain what the use of all these clouded constructions can be in present day analysis of law i.e. the post-modern law of a rapidly changing society.

2. The Concept of „System“ in Traditional Legal Approaches and in Systems theory

When we talk about systems in the legal sciences we tend to think about systems of sentences or rules: there is a separation between the part and the totality - general and specific sentences - the latter being characterized by separate 'competencies' and differentiated applicability which avoids overlaps - to name but a few characteristics of a legal system in the traditional sense. With reference to law, we presuppose in our approach the unity of a system which, in the continental tradition, allows for a deductive conception of the application of a legal text to a specific case. Naturally, we would no longer accept the idea that all the specific problems we are confronted with a preformulated or, at least, a predetermined solution in the system of the statute and "the" law as such.[1] And, of course, Common Law allows for a different logic proceeding in a somewhat analogical approach from decision to decision instead of a deductive application of a rule to a case. But Common Law also needs some systematic structure in order to allow for the search for the right answer in the existing body of decisions.

What we can retain as a first preliminary assumption is the idea that the concept of systems used in legal thinking tends to presuppose a predetermined meaning of the law which, of course, allows for, and even demands, creative interpretation when we are confronted with new hitherto unheard of cases. This is the point where we need - as Alexy would put it - a 'supplementary normative assumption', which has to be controlled in a rational way. This would be the point where a procedural discourse-theoretical approach based on rational argumentation about the correct solution would come to the fore. It would again propose a system of rules and principles for a discourse about the solution of new cases. As Alexy puts it[2]: This system is

1 Cf. G.Vattimo, Etica dell'interpretazione, Torino 1989, who tries to open hermeneutics to conditions of post-modernity; for a more traditional approach cf. H.G.Gadamer, Philosophical Hermeneutics, Berkeley 1977; for legal hermeneutics cf. F.Müller, Juristische Methodik, 2nd ed., Berlin 1976, id., Strukturierende Rechtslehre, 2nd ed., Berlin 1994; M.Kriele, Theorie der Rechtsgewinnung, entwickelt am Problem der Verfassungsinterpretation, 2nd ed., Berlin 1976.

2 Cf. R.Alexy, A Theory of Legal Argumentation. The Theory of Rational Discourse as a Theory of Legal Justification, Oxford 1989; id., Recht, Vernunft, Diskurs.

based on conceptual transparence, on clarity, empirical information, universalizability of the argumentation and freedom from prejudice[3]. These principles, which are not directly applicable to a case, have to structure the process of argumentation, thus aiming at the generation of a new rational solution to a case, whose definition cannot in a strict sense be derived from existing rules.

It is, namely, the open procedure which allows for the consideration of all possible arguments which should guarantee a rational use of the broader principles which do not have the character of behavioural rules but contain only partial and general preference rules; in other words, rules of balancing (not excluding but to the contrary presupposing the possibility of conflicting principles and priorities).

Alexy presents this advanced systems-concept as a model with three layers or levels: rules, principles and procedures. All of these are considered to permit an answer to the question of what the legal system tells us about which decision to make in a specific situation. Even this evolutionary and proceduralized approach presupposes a unity of the legal system based on the idea of universal practical rationality and a universalistic practice of decision-making. In this sense, it still follows the tradition of modern rule-based assumptions about rationality. Beyond simple cases of application of rules, the universality of the law has to be guaranteed by principles on the formulation and inclusion of new rules into the system. Under conditions of complexity[4], decision-making can only derive legitimacy from the procedural requirement of the inclusion of all the arguments to be considered. This argumentative constructive procedure follows the model of the universal rationality of the law as a system of rules - both of which are based on a presupposed fundamental basis guaranteeing the consistency of sentences[5].

Studien zur Rechtsphilosophie, Frankfurt/M. 1995: id., Law, Discourse, and Time, in: Archiv für Rechts- und Sozialphlosophie, Beiheft 64, 1995, 4ss.

3 Cf. J.Habermas, Faktizität und Geltung, Frankfurt/M.1992.

4 Cf. for the interpretation of "complexity" H.Atlan, Entre le cristal et la fumée, Paris 1979; J.P.Dupuy/P.Dumouchel (eds.), L'auto-organisation. De la physique au politique, Paris 1983; H.Maturana/F.J.Varela, Autopoiesis and Cognition, Boston/Dordrecht 1980.

5 Cf. K.Larenz/C.W.Canaris, Methodenlehre der Rechtswissenschaft, 3rd ed., Munich 1995.

The approach draws on a concept of unity of the law which is based on a general understanding of a potentially transparent reality. This reality may not be accessible to a rule-like order kept separate from the process of application in the traditional sense. However, rational discourse allows for a reconstruction of the universality of the law through the process of argumentation which is characterized by the willingness of participants to overcome the limitations of views narrowed by the practical constraints created by the concrete situation or the specific interests (of the parties to a contract, etc.). The ideal observation is no longer guaranteed by the presupposition of a world which is accessible to analytical distinction of properties which may be used in order to structure the flow of reality and allow for generalization of rule-like linkages and stable expectations of continuity on which universal rules can be based. But argumentation may create a functional equivalent for rules in the form of a rational superimposition of both a general reflection on concrete situations and the arguments to be considered which would have to pass a procedural test in order not to be linked to private (limited) interests.

3. What is Autopoiesis?

A. "Closure" of the System - Animadversions on the Concept of 'Autonomy'

The autopoietic concept of the system in general - we will come back to the legal system later - is based on the assumption of a necessary continuation, a self-production of a system[6] - not on some essential unity to be presupposed

[6] Cf. N.Luhmann, Soziale Systeme, 4th ed., Frankfurt/M. 1991; id., The Self-reproduction of Law and its Limits, in: G.Teubner (ed.). Dilemmas of Law in the Welfare State, Berlin/New York 1985, p.111ss., id., The Unity of the Legal system in: G.Teubner (ed.), Autopoietic Law. A New Theory of Law and Society, Berlin/New York 1988, p.12, 20, 32; G.Teubner, Law as an Autopoietic System, Oxford 1993; H.Baxter, Autopoiesis and the 'Relative Autonomy' of the Law, in:

and reproduced - but on a process linking event to event - operation to operation - in a differential mode. It does not proceed in a top-down approach sticking to the preservation of a settled core of rules but, instead, is rather in search of network-like patterns of linkages to be generated in a bottom-up mode leading to the construction of a distributed order linked to its own process of self-generation. This would allow for a first characterization: this idea of a system is heterarchical, not hierarchical, i.e. it does not presuppose a fundamental unity or a universal rationality or the like. It generates rules which cannot be kept separate from the process of application. It does not accept the presupposition of an ideal observer who should be able to refer to a rationality which is not itself linked to the process of the self-production of the system. The focus is on the search structure - the self-production of order within a system. The system-character of this process of self-production, linking operation to operation - in the case of the legal system, this would be decisions, contracts, etc., - would consist of a functional closure establishing a self-limitation - a selectivity of linkages of the production process: as soon as there is self-organization of the reproduction of the linkages between certain operations, other operations following different patterns of combination, which may be sorted out (for example, economic operations) are excluded from being used as a part of the system's process of reproduction. A system constructs order by making differences between itself and its environment (which, to a large extent, is composed of systems itself) and it uses its own history of operations in order to find orientation for the linkage between operations and the further retention of patterns to be reused within the continuing process of self-construction[7].

The identity of the system consists of a recursive closure which allows for the reproduction from operation to operation[8]. It is a kind of self-construction or self-constitution referring to a certain 'Eigenvalue' which is neither a fundamental rationality, nor a goal, procedure, or a concept of rationality but

---

Cardozo Law Review 1998, 1987, 2003; J.Clam, Droit et société chez Niklas Luhmann. La contingence des normes, Paris 1997.

7 Cf. Clam, supra, note 6, p.261.

8 Cf. N.Luhmann, Openness and Closure: On Reality in the World of Law, in: Teubner (ed.), supra, note 6, Autopoietic..., p.335, 345; id., Some Problems with "Reflexive Law", in: G.Teubner /A.Febbrajo (eds.), State, Law and Economy as Autopoietic Systems, European Yearbook in the Sociology of Law, Milano 1991/92, p.389, 392.; Clam, supra, note 6, p.267.

the set-up of a selectivity linking operations and allowing for recursivity. This is what closes the system and what makes it autonomous – again, not in the fundamental, anthropomorphic way of human autonomy. There is no absolute beginning of the system. There are no fixed rules of structures which command the process of self-preservation and self-reproduction of the system. There is no essence. The system has an emergent character, i.e. it does not exist outside its concrete operations as a set of principles, a teleological goal, etc. The system does not consist of individual or fixed components like atoms but its components, its basic elements, are events[9] - temporal links between operations. Its components are no less given than its structure. The operations and the relationship between them are retained and stabilized as structures - patterns which are reproduced and allow for the differentiated selectivity of the reproduction.

Recursive closure is not equivalent to the arbitrariness of connections, as connections are selective; this means past and future operations reduce and limit the complexity of possible relationships between operations. Thus, the system cannot operate on the basis of a description given from outside - a „correct" description of reality, it does not have access to. Closure is not to be confounded with impermeability to influence from outside: the system is permanently "irritated" from outside, i.e. by its natural unstructured environment or by other systems. But systems do not share a common reality - they have to and are only able to observe and operate on certain elements of reality - because their own "identity" does not have a stable objective character, they are a product of their own operations using certain distinctions and neglecting others. Systems do not have a transcendental metaphysical identity as was attributed to the subject as the bearer of the essential law-like structure of the world beyond the fragmented character of empirical reality. That is why systems are said to be linked to a "polycontextural" conception of reality: "reality" is not a fiction, however, it is a mixtum compositum of different practical constructions produced by law, politics, economics, etc. which can be tested on the basis of their success which is evaluated according to internal distinctions.

9 Cf. N.Luhmann, Die Geltung des Rechts, in: Rechtstheorie 1991, 273, 280; Clam, supra, note 6, p.249.

B. *Systems as 'Historical Machines' Operating with Distinctions*

This mechanism of recursive reproduction - i.e. the effective continuation of the system is linked to a binary code[10] - in the case of the legal system, the distinction between law and non-law - which is its guiding distinction that keeps the system and its recursivity in eternal movement: each operation opens new connections - constraints and possibilities. The retention of structure, of memory, so to speak, of successful operations opening new relations does not have a teleological function - it is connected to its own processing. This is a kind of 'machine' - even a 'historical machine' - with a memory - but a memory which is based on distinctions with no direct access to a holistic view of reality.

One could use the metapher of a blind man using a stick to test the stability of the ground on which he walks. He draws the distinction stable - unstable and constructs a whole recursive system of orientation on the basis of this chain of operations which allows him to walk but does not permit him a full description of his environment. The paradox of closure can also be demonstrated with reference to this example: if the blind man realizes that his stick and the operations he performs allow for differentiation on a specific sensibility, he is able to set up quite a complex construction of his surrounding. It is the closure of this system - I shall come back to the theoretical questions - which allows for openness. This paradox can and must be de-paradoxified: if the system reduces itself to the organization of its operations (executed by the stick), it can find a productive way of coupling to its environment. The extreme and unstructured complexity is only accessible to the system if it develops a 'translation'-system of its own which does not correspond to its external reality. This is the explanation for the code - it is so to speak the blind man's walking stick.

---

10 Cf. N.Luhmann, Das Recht der Gesellschaft, Frankfurt/M. 1993, p.69ss.

C. Systems only Exist in the Plural Form

This is why, however, the system needs an environment which is composed and structured by other systems. Systems always exist in the plural form: there are different systems, all of which operate on their own 'Eigenvalue' and allow other systems to use pre-structured complexity. For example, for the legal system it is absolutely necessary that other systems, science, economy, politics, etc. also develop their own autonomy - otherwise the autonomy will break down. Again, the example of the blind man can be helpful: he has to presuppose that there is a pre-structured order separating pavements from roads, that people in general are polite and not rude, distribution of shops, etc.. Otherwise, orientation would be impossible because of unstructured and inaccessible complexity.

The system is a kind of self-creating network of relationships which designs itself on the basis of linkages which have already been operated successfully, i.e. which fit into the structures and the patterns which have 'worked'. The closure of the system does not mean isolation from external influences - on the contrary, the system is operationally closed which means that it is open to coupling, but only on the basis of its own operational and semantic possibilities[11].

1. The Legal System - Some Basic Theoretical Elements

A. The Guiding Distinction of the Legal System

The legal system has to observe the economic system but only on the basis of its own distinctions, e.g. the legal system has to adapt property to a pre-structured 'reality' created by the economic system and vice-versa. The

11 Cf.Luhmann, supra, note 6 (The Self-reproduction...), p.113; Baxter, supra, note 6, 2003s.

distinction the law operates on and which allows for its self-production through legal communication is linked to the function of the legal system - a specific function which consists of the counter-factual stabilization of expectation[12].

The function of the law is separated from the function of morality, which operates on a basic good/bad distinction. This differentiation produces the side-effect of blocking the immediate influence of vague feelings of justice, for example. The law does not pretend to have access to 'justice' as such. However, every approach to the establishment of a meta-norm as a stable source of validity runs counter to the necessary function of a positive validity which is self-generated by the system. But, by its operational recursivity, it destabilizes any principle of validity - it includes decisional instability into the operating mode of the system itself because it excludes any axiological reference to fundamental values and principles. The architecture of the system implies a limit to the centralization of the legal material because it uses a differentiated search and test structure and excludes moral disqualification of persons as incompatible with systemic differentiation in general, and, in particular, the set-up of a sophisticated internal structure of the system. Such a system cannot work on unstructured values or moral discriminations. It might only work on distinctions elaborated by its own memory. Of course, nobody can exclude that the autonomy of the system, including the legal system, would be destroyed or damaged, for instance, in a nazi or communist régime. But a system as such would instead reject this type of intrusion as being incompatible with its differentiation.

B. The Distinction of Code/Programme and Centre/Periphery in the Legal System

We have already increasingly referred to the legal system as an example of autonomous systems in general. We should now focus a bit more on some supplementary characteristics of the law: the code of the legal system alone

---

12 C.f. N.Luhmann, Ausdifferenzierung des Rechts. Beiträge zur Rechtssoziologie und Rechtstheorie, Frankfurt/M. 1981, p.73ss.; id., Rechtssoziologie, 3rd ed., Opladen 1987, p.43, 50; Clam, supra, note 6, p.132.

would not allow for the set-up of structure within the system. The system needs programmes[13], e.g. a legal act for its functioning which specifies its distinctions between law and non-law and allows for the generation of patterns, such as legal dogmatics. At this point, a supplementary transformation of the systems approach has to be introduced, which again runs counter to the hitherto established assumptions: according to N. Luhmann, the judicial decision-making system is the centre of the legal system whereas the legislature is to be located at its periphery[14]. The reason why is again to be located in the machine-like character of the legal system: it produces decisions and reinforces its autonomy on the basis of repeatedly creating new relationships and new constraints which, in turn, open new connections. And this pressure is mainly imposed on judges, and not on Parliament as the legislative body.

1. From a Universalistic to a Relational Paradigm of Law

A. What is the Basis of Universalistic Law?

I would now like to make some remarks to try to explain the rationale of this construction - which may appear terribly counter-intuitive. At this point, we should perhaps take a step back to R. Alexy and his emphasis on a substantive concept of justice and a procedure of rational argumentation supplementing the weakened rationality of the liberal concept of universal law which separates the specific and individual conditions of people and their statuses from an abstract universalisable rule of collective order. It is this conception of accessibility of a hyper-complex world to a substantive stable concept of practical rationality which is at stake. This assumption of a universal stable rationality of the legal system, of justice, of a collective order and of the individual as its basic element - i.e. the conception of the enlightenment - even in its more flexible versions, loses its force once we can

13 Cf. Luhmann, supra, note 10, p.69ss.; id., Die Codierung des Rechtssystems, in: Rechtstheorie 1986, 171ss; Clam, supra, note 6, p.281.

14 Cf. Luhmann, supra, note 10, p.333.

no longer talk of the single integrated reality to be referred to a stable framework of rules of human thinking penetrating the essence of the world.

Society has lost its centre, and the individual can no longer be regarded as the basic component of society. Rather - and at this point the systemic construction tries to take up the new challenge - the complexity of the relationships between the individuals has augmented and gone beyond a point of no return: this means they are no longer to be integrated into a common shared understanding of reality, and this was exactly the basic assumption underlying the concept of the law-like character of rationality and the universality of rules: there is a complexity of individual and specific relations which need not, and cannot be, taken into account although we can have access to the structure, the reproduction of patterns of behaviour and of collective order once we know the universalizable rules, i.e. of substantive rationality. No longer is there a stable point from which the whole of the society can be observed and, moreover, this is not even an acceptable idealization.

B. *Why Relationships Matter*

If we accept this as a provisional hypothesis, we may accept the rather strange assumption of Luhmann's systems theory that society is not composed of individuals but of communications. This idea is not as strange as it seems: if we take a comparative look back at universal rationality, we could come to the conclusion that this conception is based on the subject only as a counterpart to universal rules, as the conscious bearer of law-like stable rationality. And once the stable law-like repetitive character of rationality is undermined it could also be acceptable to focus on the relationships between the individuals and look at the patterns of change which may be observed and may be the object of stabilization and intervention and, finally, could be used for the construction of a new heterarchical relational rationality. Such a rationality would not be accessible from a stable point of observation. But we would have to observe that this type of reality, of a world consisting of overlapping networks of interrelationships, is only accessible to the polycontextural observations generated by the differentiated systems of society. A relational concept of rationality could then be based on the search

for comptability between different systemic logics of law, economics etc.. There is no room for a unitary approach to reality which could be reconstructed in the rule-oriented reflection of a knowing subject and this unity cannot be guaranteed by an intersubjective process of self-reflection in argumentative processes, either[15].

Rationality, as an ordering principle can, nowadays, only be referred to trans-subjective systems of communication being generated by networks of inter-relationships[16]. These connections continuously create both constraints and options for new legal communications. This new type of sustainable relational patterns replaces the rule-based structure of universal rationality which can be mastered by the subject and its rational conscience and reflection. The new legal perspective is no longer centred on a body of rules and the rationality of its author's will which are taken to be separated from the practice of legal communications which, in turn, are regarded as its „application". This practice, however, does not have major repercussions on the rules, which are themselves derived from rational universalistic reflection. In N.Luhmann's view[17], it is judicial practice, instead, which is the core component of the legal system - and not laws and rules themselves. This does not mean that law simply does not underlie any influence from the political system (Parliament). Of course, laws are a crucial element of the legal system, but, in a longer term perspective, this is only the case if judges are able to transform them into practical decisions; and, if they do not adopt legal „systematicity" (accepting specific legal constraints), this will be difficult or even impossible. For this reason, the legislator has to consider the relational patterns which are reproduced in the legal practice. I would even go one step further and venture the hypothesis that not only judicial decisions but also, and even primarily, individual legal acts (contracts, etc.) have to be placed at the centre of the legal system themselves.

---

15 Cf. generally N.Goodman, Fact, Fiction, and Forecast, Cambridge(MA) 1983; H.v.Foerster, Observing Systems, Seaside (CA) 1981; id., Sicht und Einsicht, Braunschweig 1985.

16 Cf. W.Welsch, Vernunft. Die zeitgenössische Vernunftkritik und das Konzept der transversalen Vernunft, Frankfurt/M. 1996.

17 Cf. Luhmann, supra, note 10, p. 333, id., Die Stellung der Gerichte im Rechtssystem, in: Rechtstheorie 1991, 273ss.; Baxter, supra, note 6, 2017.

N. Luhmann would not have accepted this idea because in his view contracts and other legal acts are not so tightly linked through connection constraints as judicial acts are. But, to my mind, this assumption does not fully exploit the advantages of a network-based legal theory. It is also private legal practice which maintains the productivity of the pool of variety within the „population of (legal)ideas"(viable forms of contracts, legal experience, trust etc.). This is all the more important because, especially in a society which undergoes a continuous process of change, cooperation is much more essential for the viability of a legal system, and not judicial interpretations and sanctions.

One of the strengths of a systems theoretical approach to law consists of the openness for legal pluralism which will be welcome once the link between law and the will of the state as its author is called into question.

1. The Challenge of Legal Pluralism

A. Standards, the New Law Merchant and the Globalization of Law

From these reflections, a perspective on the new law merchant can be derived[18]: if legal methodology was able to get rid of the orientation on state-based decision-making and stress the potential of the „bindingness" of self-generated relationships between legal communications, their observation and evaluation, it should be possible to accept the new law merchant as a form of law like any other and not to look for a public mechanism of transformation into state based law (by way of the decision of a judge, for example). The same could be valid for private standard setting (CEN, ISO and similar

[18] Cf. G.Teubner, The Global Bukowina. Legal Pluralism in the World Society, in: id.(ed.), Global Law without a State, Aldershot 1996, p.3ss.

Autopoiesis of Income Tax Law: a Problem for Reform
Draft and Notes 2004
John Prebble

Introduction

The theory that law may best be understood as an autopoietic system has gained considerable ground since Luhmann advanced it in 1981.[1] Nevertheless, there have been few endeavours to apply the theory in any practical way or to employ it to analyze particular areas of law. This article proposes a hypothesis that is improbable on the face of it: that Luhmann's theories usefully illuminate the unpromising field of income tax law. Nerhot summarizes the autopoietic theory thus:[2]

> When it is said that the law can be regarded as an autopoietic system, what that essentially means is (and this is the major departure from traditional systems or cybernetics) that it is the law itself, regarded as a system, that says what is law and what is not, with the theory defining the way in which the statement is made.

As Luhmann puts it, "Law is then valid by virtue of decisions which make it valid. The legal system itself is obliged to believe in this ground of its validity".[3] In other words, it is immanent within the law that it relies for its validity on itself and not on exogenous factors.

Autopoiesis as a jurisprudential study

Autopoiesis locates itself within the last of the four great branches of jurisprudence: sources of law, legal concepts, legal analysis, and legal theory.[4] Like most theories of law, it is a reaction to what has gone before, an attempt to construct

---

[1] Niklas Luhmann, "Evolution des Rechts" in N Luhmann, *Ausdifferenzierung des Rechtsystems. Beitrage zur Rechtssoziologie und Rechstheorie.* (1981 Frankfurt a M: Suhrcamp); "Autopoiesis, Handlung und Kommunikative Verstandigung" (1982) 11 Zeitschrift fur Soziologie 366. For a collection that develops and analyses Luhmann's theories, see Gunther Teubner, *Autopoietic Law: a New Approach to Law and Society* (1987 de Gruyter, Berlin, New York). This volume has extensive bibliographies, including listings of Luhmann's work in English.

[2] Patrick Nerhot "The Fact of Law" in Gunther Teubner, *Autopoietic Law: a New Approach to Law and Society* (1987 de Gruyter, Berlin, New York), 312, 312.

[3] Niklas Luhmann, unpublished conference paper, quoted in Jean-Pierre Dupuy, "On the Supposed Closure of Normative Systems", in Gunther Teubner, *Autopoietic Law: a New Approach to Law and Society* (1987 de Gruyter, Berlin, New York), 51, 68.

[4] See, eg, the organization of RWM Dias, *Jurisprudence* (5th ed 1985, Butterworths, London).

an epistemological explanation for law that is coherent and that resists attempts at falsification.

Most of the history of jurisprudence has been dominated by one version or another of natural law theory, that is, that laws that govern humans are derived from (and if not, should be derived from) the law of nature, being the legal and moral sense that comes from God. In the nineteenth century. philosophers rejected the alleged connection between law on one hand and morality and religion on the other. This rejection led to the rise of positivism, which, in its original form as propounded by Austin, saw law as depending on the command of a sovereign. This explanation could not account for all of the phenomena that we recognize as law; so Kelsen developed a kind of idealized positivism, known as the "pure theory" of law. The pure theory saw law as essentially independent of other social phenomena, as a hierarchy of norms or commands, each level of command inferior to and dependent on the level that precedes it. Kelsen's explanation breaks down at the last stage: what does the highest norm of a legal system depend on? Kelsen answered this question by postulating a fictional "grundnorm" that society more or less accepts: a rule that says, essentially, "Law is to be obeyed".

From one point of view, autopoiesis is a reaction to Kelsen, an attempt to explain law without recourse to fictions. Kelsen's system was nearly autonomous, in that each norm depended on a higher norm and not on something outside the law, but in the end the system is not closed and self-sufficient, because it must refer outside itself, either to the grundnorm, or to whatever fact it is that the fiction of the grundnorm obscures. In contrast, autopoiesis postulates that law is a closed system of norms that are not hierarchical but circular.

This brief description shows autopoietic theory to be squarely within the tradition of post-modernism, (if it is not a contradiction in terms to speak of the post-modern as inhabiting a "tradition"). Using Norris's somewhat unsympathetic description, post-modernism is characterized by:

> ... a laid-back pluralism of styles and a vague desire to have done with the pretensions of high modernist culture. In philosophical terms, post-modernism shares something with the critique of Enlightenment values and truth-claims mounted by thinkers of a liberal-communitarian persuasion; also with neo-pragmatists like Richard Rorty, who welcome the end of philosophy's presumptive role as a privileged, truth-telling discourse. [Another

> characteristic is] the current preoccupation, among some philosophers, with themes of "self-reflexivity", or the puzzles induced by allowing language to become the object of its own scrutiny in a kind of dizzying rhetorical regress.

In terms of this definition one may think of Kelsen's pure theory as an apogee of Enlightenment thinking, with theorists turning to autopoiesis as a reaction against what they perceive as ultimately sterile.

Origin in biology

To be added

Difficulty of applying autopoiesis theory to most law

Company law

Clearly develops in response to a need. People want to enable the divorce the operation of businesses from the provision of capital, so that businesses can attract the investment that they need and so that investors can have structures that enable them to provide capital for ventures that they do not have the time or the skills to take part in. Also, allows many small capitalists to aggregate their small investments so that there is enough capital for large projects.
Original response of governments: royal charter (often accompanied by a monopoly); individual Acts of Parliament.
Makes sense to legislate for a system of incorporation by registration: Companies Acts.
Companies Acts steadily reformed and developed to meet the needs of business: directors' duties and liabilities, floating charges, protection of minorities, prevention of monopolies from exploiting position (compulsory purchase)
In short, does not look like a closed system.

Contract

Forms of action (somewhat closed system)
Slade's case: requirements of increasingly commercial society impinge, break into the self-generating system.
Perhaps some aspects are autopoietic. Postal acceptance rule? No. Based on practicality and fairness: who should bear the risk?

Income tax law: perhaps a lone example of autopoiesis

Form over substance
Fictions

Teubner theory

Initially, legal systems are open. People construct them to create rules to govern social facts. But once they reach a certain level of maturity they replicate themselves, becoming autopoietic.
Prebble response. Not really so. Law reformers are always taking account of social facts and changes in society, eg gay marriage.
But income tax law may be autopoietic in Teubner's sense. Clearly so in respect of judicial doctrines: capital-revenue, timing. Also so in respect of some law reform. Eg, any tax preference; IRD consultative document on recruitment of foreign workers, 2003. Does not follow principle. Pragmatic, intra-tax ideas (though motivated by social facts).
Another aspect: base broadening/equity measures inevitably lead to further and further refinements. Not so in the same sense in other areas of law reform. Eg, depreciation.
Ashby's theorem: a programme programming itself is a logical impossibility; changes must be produced from the outside. It may be that in a sense income tax law breaches Ashby's theorem.

Between Order and Disorder: The Game of Law

FRANÇOIS OST
Bruxelles

Introduction

Radical theories generally arouse contrasting reactions. The discovery of the *galaxie auto* is no exception to this rule. The observer hesitates between fascination (Livet, 1983: 165 ff.) and consternation (Gutsatz, 1983: 29 ff.). Sometimes one has the feeling of being present at some new Copernican revolution: not just the discovery of a new scientific continent, but a radically new illumination of our previous knowledge, the emergence of a genuinely post-Cartesian *episteme* which substitutes positive operational concepts for the categories of object, subject and linear causality heretofore affected solely by the "suspicion" of modernity. Other times, instead, the predominant feeling is irritation at what may seem like a magic trick: the ideas of the "self-programming program", of "linkage through closure", of "co-evolution of autopoietic systems", once the first impression of strangeness they arouse is dissipated, lead one to think that "there must be a catch to it", some secret mechanism activated by the demiurge that presents them — which ultimately tells us more about their authors' ingenuity than about the nature of the phenomena studied.

This ambivalent reaction is aggravated still further by reflexes bound up with the academic compartmentalization of disciplines. What is one to think of concepts ("selection", "information", "communication", "organization", "autonomy", etc.) forged in the field of the social sciences, then borrowed by biologists to explain the living world, at the cost of reformulating them in the language of cybernetics, only to be at length reintroduced into the social sciences? Ought one to welcome the abolition of the narrow boundaries of academic fields and announce the arrival of a transdisciplinary science, or ought one to denounce the ideological hocus-pocus consisting in legitimating attempts at social domination using theories lent the prestige of science by their bio-cybernetic trappings?

Is it then once again to be the conflict of science and conscience, the old humanism that hesitates between celebration of the technico-scientific *logos* that manifests the increasing sovereignty of theoretical reason and

secular protest in the name of the practical reason that is the bearer of a freedom – and therefore of a history – that no system can ever contain? But what if, instead, as one's forebodings sometimes suggest, this phantasmagorical hypostasis of the "system" conceals some shameful theology, so that the scientific position ought to be more on the side of those who cling to the indeterminable, the random, the irreducibly plural?

As for ourselves, having been induced to take part in this discussion – at least as far as the explanatory power of the autopoietic paradigm in law is concerned – we shall endeavor for as long as possible to maintain the ambivalence described. Guarding ourselves, then, from either dogmatism or *a priori* skepticism, we shall enquire into the fruitfulness of the autopoietic theory, which today lays claim to being the new paradigm for the social sciences. In doing so, we shall divide what we have to say into four stages.

A first section will succinctly present the autopoietic paradigm as formulated, in biology, by Maturana and Varela. A second section will look at Luhmann's attempt to transpose these ideas into the field of law. In this connection we shall recall some theses of Kelsen, who seems to us to be a thinker of the autonomy, if not the autopoiesis, of the legal system well before the bio-cybernetic formulation of these concepts.

We shall then devote a third part to the critical discussion of the self-referential model. We shall take our objections both from cybernetic theory (Ashby), biology (Atlan) and philosophy (Castoriadis). We shall then be in a position to discuss, in a fourth section, a substitute paradigm, that of the game. This seems to us to have the advantage of retaining many of the ideas of the autopoietic theory (autonomy of the game, internal regulation, etc.) while balancing them through taking account of the player, whose strategies, though obviously not purely subjective and random, cannot for all that be reduced to implementation of a pre-established program. It is this intermediate space between the objectivity of a self-regulated system and the subjectivity of the player's action – the space of the game, the locus of transformation of the project into system and the system into history, the change from disorder to order and from order to disorder – that seems to us to be simultaneously constitutive of the social and capable of being better illuminated by the game model than by that of autopoiesis.

I. The Paradigm of Autopoiesis

It is no doubt premature to accept the existence of *one* paradigm of autopoiesis. From various quarters, in fact, theories on the organization of life have been worked out; but the relationship between these concepts applied certainly does not yet authorize us to speak of a single paradigm. We shall, therefore, to simplify, content ourself with bringing out the ideas

of Maturana and Varela, particularly since it is to these that Luhmann principally refers.

The approach to life proposed by Maturana and Varela is deliberately mechanistic: living systems, like the cell or the brain, are to be understood as machines explained by their internal organization (Maturana and Varela, 1980: 75; Varela, 1979: 4). Any idea of force, intention, teleonomy, adaptive transformation or, generally, any principle that cannot be found in the physical universe has to do with an animist perspective, with an artefact of language that tells us nothing about the organization of life (Maturana and Varela, 1980: 85). Living systems, like machines, are goalless systems. But while in the case of an allopoietic machine, such as a car, the product of its operation is different from itself; in contrast the product of the operation of an autopoietic machine is nothing other than itself (Maturana and Varela, 1980: 89). The cell can thus be defined by its self-generation, on the model of an artificial machine constructing itself.

An autopoietic system can then be defined as a machine organized as a network of processes for producing components which, by their continual interactions and transformations, unceasingly regenerate the network of processes for producing components and thereby give the machine a definite spatial unity (Maturana and Varela, 1980: 89; Varela, 1979: 17). The autopoietic machine, then, is a network of tangled or "looped" processes that ensures constant regeneration and specific unity in space. Autopoiesis is both the necessary and the sufficient condition for defining life. Several consequences follow from this definition.

First of all, an autopoietic machine is autonomous. This autonomy must be understood as the capacity to subordinate all change to the maintenance of self-organization. Autonomy here links up with homeostasis: not inertia or stasis, like that of a crystal for instance, but the constantly re-established equilibrium of a network of processes in interaction. It will then be noted that the autopoietic system has an individuality: to the extent that it keeps its identity constant through a process of continual adjustment, this identity is in some sense self-produced and owes nothing to any intervention by an outside observer. An autopoietic machine itself defines its boundaries by determining what is and what is not itself.

Finally, the conclusion will be reached that an autopoietic system has no inputs or outputs. Certainly, it may undergo exogenous shocks, independent events or "perturbations"; but these, far from acting as information that would contribute to programming the system, arouse internal compensating reactions, so that the homeostatic equilibrium characterizing the system is kept invariable. As long as it controls its processes of transformation by subordinating the amount of change to the maintenance of its organization, the system will preserve its dynamic unity; it will remain autopoietic (Maturana and Varela, 1980: 80).

This last feature is no doubt the most significant one. It entails, according to Varela, a radical paradigm shift. Whereas classical scientific thought holds to a paradigm of "linkage through inputs" (the outside determines the changes in the system), what is being proposed here is a paradigm of "linkage through closure": it is the internal coherence of the system that determines its development. While in the first case it is the environment that supplies the fixed points that enable the evolution of the system to be understood, in the second model its evolution is determined by its self-organization, while the points of contact between the system and its environment are reduced to mere "perturbations".

The major concept in this new perspective becomes that of operational closure. An autopoietic system is an operationally closed system within which dynamic coherence engenders a wider or narrower variety of eigen-behavior. Thus, the "history" of a species should not be understood as optimization of its adaptation to the environment (which would presuppose a logic of "linkage through inputs"), but instead as the diversification or complexification of forms of eigenbehavior authorized by the coherence and operational closure of the system, which thus manages to maintain its identity through its transformations (Varela, 1983: 149–53).

Autopoietic systems, then, are not isolated: they define themselves against the background of an environment that is a source of exogenous shocks, and they can enter into relationships with other autopoietic entities. Varela speaks in this case of "structural coupling". Here again it is not a case of autarchy nor of outside programming; quite simply, each autopoietic system entering into a relation of structural coupling is, with respect to the others, a source of compensable perturbations. If these perturbations exceed the margin of variation tolerable by the system in order to keep its identity, there would be disintegration of one or more of the copresent autopoietic systems (Varela, 1979: 32–3; 50ff.).

Having thus briefly presented the major outlines of the autopoietic paradigm, one may now ask to what extent Maturana and Varela encourage the transfer of this model to the explanation of social organization. The least one can say is that the answer to this question is doubtful. In their jointly written work, Maturana and Varela say that they have not reached a common position on this point and therefore decide to postpone the discussion (Maturana and Varela, 1980: 118). Varela is however more explicit in his own book. For him, the concept of autopoiesis is not capable of being imported into the area of social phenomena. It is ill-advised, he writes, to use the concept to describe, for instance, an animal society: the distinctive features of its unity are not of a topological order, and the social interactions taking place in it cannot be understood in terms of "production" of components. The same is true of human systems: treating them as autopoietic proceeds from a confusion with the idea of autonomy.

For Varela, then, social systems are capable of being explained in terms of autonomy and not of autopoiesis. Autonomy does presuppose operational closure: there is the circular connection (looping) of organizational processes, which reflects an intrinsic coherence, a self-organization, but that does not mean one can speak of self-generation of components of the system (Varela, 1979: 54 ff.).

No one, however, controls the fate of the concepts he puts in circulation. (Could there be in this indeterminacy of the march of ideas a form of autopoiesis?) We must, then, look at how the autopoietic model is transposed into the social field, specifically into law.

II. The Transposition into Law of the Autopoiesis Paradigm

1. Luhmann and Teubner

The idea of analyzing social phenomena and universes of discourse in terms of autonomy or increasing autonomization constitutes an endeavor which is neither new nor illegitimate. Thus, for instance, it was using concepts of self-organization and self-finalization that Ladrière described the evolution of the scientific logos and of the technological world that accompanies it. The scientific method, a highly pondered and self-controlled specification of the rational method, tends to engender an autonomous reality subject to its own law, which is a law of growth. Thus, "within contemporary societies, science tends, under the action of its own internal operations, to constitute itself as a vast system, itself formed of interacting subsystems, evolving towards increasingly complex and increasingly integrated forms, which are at the same time increasingly autonomous" (Ladière, 1977: 51). Though not able to free itself totally from outside circumstances, scientific discourse is today capable of submitting them to critical assessment so as to arrive by successive equilibration at almost complete self-control. As for the technological superstructure, it too is inhabited by a self-organizing and self-finalizing movement: "one might say," maintains Ladrière, "that it is commanded by an immanent teleology of which it is itself the source" (1977: 70).

Luhmann and Teubner's endeavor is, however, more ambitious: it is the law itself, as a universe of discourse and an institutional field, that might be described in categories no longer simply of autonomy and self-organization, but of autopoiesis.

How can the theory of autopoiesis be conserved while stripping it of its biological connotations? This is the first question raised by Luhmann (1985 a: 2). The answer lies in bringing out the minimum constitutive element of the system — a sort of legal atom. In this case it is communications (Luhmann, supra: 16—17). More specifically, Teubner takes the legal act as the basic element of systematics in law (Teubner, infra: 221).

Accordingly, a legal system may be called autopoietic to the extent that the self-reference characterizing it affects not only its structures and organization but the basic elements themselves. This, according to Luhmann, constitutes the fundamental paradigm shift (Luhmann, 1985 b: 112). It will be noted from the start that the transposition is effected at a radical level: that of the autopoiesis of the basic constituent elements of the system and not merely that of autonomy. This move, however, seems to be repudiated by Varela himself. Luhmann is quite formal on this point: the law, like any other operational, well differentiated social subsystem, presupposes itself and reproduces itself: both its unity and its organization, its constituent elements and its boundaries, result, by "complexity reduction", from intrinsically systemic performances; they proceed neither from nature nor from any condition coming from the environment. Consequently, there is no law other than positive law (whether legislative, case law or contractual) (Luhmann, 1985 b: 112).

An autopoietic system is, then, above all a closed system. In law, this closure is normative. Only legal norms can decide on the pertinence or relevance to the legal system of any element whatever. After centuries of discussion, it is today admitted that neither nature nor morals nor religion have this power to create law. Only the legal system, through its autopoietic functioning, is capable of conferring the quality of legality on the elements that it determines. Thus, an infinite and non-intentional process is brought about, of reproduction of legal elements by themselves (Luhmann, 1985 b: 113; supra: 20).

But the advance of the autopoietic paradigm consists in going beyond the classical opposition between openness and closure which previously affected systems theory. An autopoietic system, defined as functionally (normatively in the case of law) closed, is also a system open to the environment, a source of information for it. The normative closure is thus parallelled by cognitive openness. A legal system can be programmed in such a way as to render itself deliberately dependent on the evolution of outside circumstances; a change in the program itself under pressure from the environment is likewise possible (Luhmann, 1985 b: 113; supra: 20—21), as long as the system does not in this search for optimum adaptation lose control over its transformations. On this point, Luhmann admits that there is a difference between him and Varela, who, as will be recalled, rejected the idea of adaptation of the system to the environment (Luhmann, 1985 b: 122).

In combining closure and openness in this way, the legal system cannot lay claim to complete self-determination. Thus, no legal system can wish to determine in advance the content of any legal decision. It suffices that the act of legal decision-making be normatively referred to a need to perpetuate the self-production of the legal system — it may, moreover, incorporate all outside information necessary to make it more functional. It will be noted in this set of ideas that the majority of legal norms are conceived of as programs

aimed at the normative treatment of information from outside ("if such and such a fact has occurred, then such and such a normative consequence shall be applied") (Luhmann, 1985 a: 28; 1985 b: 116 ff.).

By now it will be clear: in order to ensure its autopoietic development, it is enough for the legal system to control a binary logic — the logic of permission and prohibition, of legal and illegal. Because they manipulate this code of identities and differences, legal systems may well open up to the outside without losing their identity, since the reality so produced — and imposed, let us say — is in fact "nothing but the correlate of a self-referential mode of operation which reproduces itself according to this code" (Luhmann, supra: 16).

The idea likewise emerges that no transcendence imbues legal systems; if it is true that these systems function like cybernetic machines, using a code of differences, there is then no need for a hierarchy of norms, a regulating principle or a fundamental norm. The autopoietic process is perfectly symmetrical, circular and continuous (Luhmann speaks in this connection of *"creatio continua"*): each element receives its normative quality from another element which it in turn determines, without it being possible to discern in this self-referential circuit any priority or primacy whatsoever (Luhmann, supra: 20—22).

Let us add one last remark of an epistomological nature, which moreover follows directly from the theses presented: the analysis of autopoietic systems proceeds from a form of "self observation"; it involves an internal analysis (Luhmann, 1985 b: 125).

Teubner for his part puts forward some complementary theses. For him, legal systems become progressively autopoietic when, at the end of a process of autonomization, they finish up by producing their own elements, the legal acts which in turn reproduce the system of production (Teubner, infra: 221—223). The evolution traced out in this way is interesting, but it may be wondered to what extent it is compatible with the analyses of the authors of the autopoiesis paradigm. Varela at any rate asserts that the establishment of an autopoietic system is not gradual: a system either is autopoietic or is not (Varela, 1979: 27).

It is nevertheless the case that, for Teubner, a system may be called autopoietic when the three characteristic functions of the process of evolution (variation — selection — conservation) are "internalized", regulated by mechanisms intrinsic to the system and no longer commanded from outside. Whereas heretofore social norms determined the possibilities of variation of legal norms, selection among the latter was brought about by phenomena of social approval and recognition, and conservation of norms adopted was assured by the dominant ideologies and myths in the society, today these three functions respond to internal regulation: at most, they are "modulated" by pressure from the environment. Thus, a normative change will be brought about only if it is structurally determined and

legally relevant; selection is henceforth effected in conformity with internal, self-produced expectations, while the function of retention is a matter of legal doctrine (Teubner, infra: 230–235).

Nevertheless, contrary to what might be suggested by these initial analyses, the stress here too is placed on the openness of autopoietic systems. These in fact enter into a relation of "co-evolution" with other systems organized on the basis of similar principles. Thus, they may mutually influence each other – at least in an indirect, mediated way. It will, for instance, be noted that the expectations formulated within the framework of a subsystem end up by adapting themselves in the long term to those prevailing in another subsystem: does the liberal anthropology underlying the category of legal subject not fit in with the market image dominating the field of political economy (Teubner, infra: 234–235)?

There would doubtless be much to be said in connection with these analyses, carried out, like those of Luhmann, from an exclusively "internal" viewpoint. Thus, it is hard to see how the idea of internal regulation of the evolution of the system and that of co-evolution can be brought together. Is it the ultra classical idea of "relative autonomy" of various social bodies? Or is it, as we tend to think, a more radical analysis that is being put forward? But how then can one, without playing with words, combine closure and openness?

Before venturing on a few critical considerations, we must recall certain theses of Kelsen, who seems to us in many respects a precursor of the ideas today being discussed under the heading of autopoiesis.

2. Kelsen: A Precursor?

Far be it from us to imagine that Kelsen anticipated present biological theories and their transposition into the field of law, which he would undoubtedly have condemned as "syncretism of methods" (Kelsen, 1967: 2). We shall content outselves with maintaining that Kelsen's conception of legal systematics is at many points very close to the view today being developed by way of autopoiesis.

The basic element of the legal phenomenon and of the science which studies it is, for Kelsen, the human act, involving on the one hand a material aspect perceptible to the senses, and on the other a meaning which is called, specifically, legal. While the meaning given to the act by its author is subjective, its objective meaning is conferred on it by the law: it is the meaning that the act has in law, according to the law. More exactly: it is legal norms, acting as interpretive schemes, that impress on certain human acts their nature as legal acts or illegal acts. But, Kelsen adds, "it will be noted that these norms are themselves created by legal acts, and that in turn these acts receive their legal meaning from other norms" (Kelsen, 1967: 3–5). It therefore seems clear that the legal phenomenon is constituted by

entities with a special meaning – a normative meaning of a legal nature – and that the latter is produced by system-forming norms, which themselves result from lawmaking acts.

There thus begins a dynamic, uninterrupted process of creation of law by itself. The following proposition from the *Pure Theory of Law* is famous: "the law itself regulates its own creation" (Kelsen, 1967: 299), ... a process "in which the law, so to speak, unceasingly creates itself" (1967: 318). This self-creation of law, albeit uninterrupted, is not thereby circular – no doubt a point of difference with the analyses in the autopoietic systems. For Kelsen, legal norms are not on the same level, but form "a building of several superposed stories" (1967: 299), a pyramid which, as we know, culminates in the supposed fundamental norm. At each stage of construction, the higher norm determines and founds the validity of the lower norm, just as a higher authority empowers a lower one. The order presiding over the implementation of the legal system is thus strictly linear and hierarchicalized, necessarily ending with the fundamental norm empowering the supreme authority to create law.

This fundamental norm confers on the legal system both the principle differentiating it from whatever is not it, its principle of unity and the foundation of its validity. It is solely because it has been created in accordance with a higher norm, and ultimately in conformity with the fundamental norm, that any legal norm whatever may be regarded both as valid and as belonging to this system. Conversely, the unity and validity of a plurality of diverse norms derives exclusively from the fact that they proceed from a single fundamental norm (Kelsen, 1967: 256).

Subject only to the "transcendent" and thus partly extra systemic nature of the fundamental norm, it will be agreed that the *Pure Theory of Law* is not far at all from the theories of self-organization – did Kelsen not take upon himself as his methodological principle to "bring together everything that is law into a system, that is, grasp it from one single viewpoint as constituting a whole closed upon itself" (1967: 430)?

But if legal systems are autonomous, that does not necessarily mean they are autarchical or tautological. The higher norm never forms more than a framework open to various possibilities, all equally regular, among which it is for the implementing body to choose, according to an act of will inspired by extra-legal considerations (Kelsen, 1967: 460).

This partial indeterminacy, which ensures the dynamism of the legal order, does not however entail loss of control by the law over itself. The legal system, according to Kelsen, has neither lacunae ("everything not legally forbidden is legally permitted") (1967: 326), nor antinomies ("a norm contrary to the norms is a contradiction in terms; such a norm would not be valid ... would not be a legal norm at all") (1967: 355).

It will also be noted that the normative interpretation for which the fundamental norm is the basis extends to every element contributing to

legal production. Thus, a collectivity founded by a legal order cannot be distinguished from that order itself (in relatively centralized legal orders that collectivity takes the form of the State; but there is a rigorous identity between State and law) (Kelsen, 1967: 379); likewise, any individual — even a mere private person — who performs a legal act is operating as an organ of the legal system, and the act he performs, a contract or will for example, can be understood as an act of the State, an act attributable to the unity of the legal order (Kelsen, 1967: 200; 373). Even an illegal act ought not to be rejected from the area of law; to the extent that it allows application of a sanction, it belongs instead to the law, as its condition and not as its negation (Kelsen, 1967: 152—6).

Kelsen does not fail to draw the conclusion from this panlegalism — which once again results from the fact that the viewpoint adopted is strictly "internal" to the object studied: "the law is like King Midas. Whatever he touched was immediately changed into gold; likewise, everything the law has to do with becomes law" (1967: 369). Or again: "in the eyes of the theory of positive law, only law can be a source of law" (1967: 313).

Each of the points raised by Kelsen is clearly capable of discussion. At this point in our reflections, we shall content ourselves with a single observation. Kelsen, it is often forgotten, explicitly pursued throughout his work a single design: to elaborate a science of law which would "advance the development from the seeds already present in 19th century positivist legal science" (Ost and Lenoble, 1980: 467—546; Ost and van de Kerchove, 1985 a: 285—324). In other words, Kelsen consistently drew the conclusions that emerge from the postulates of legal dogmatics. In so doing, he doubtless contributed more towards rationalizing jurists' discourse than to advancing a genuine science of law that explains its object (Kelsen, 1967: viii). In a sense, Kelsen is thereby expressing the truth of law: let us understand the truth of the dogmas, beliefs and myths shared by jurists, and especially their ideals of logicality, their desire for unity, clarity and security; but in another sense, the *Pure Theory of Law* disguises this truth by presenting it in a rationalized, "purged" form that conceals both its ethical extensions and its societal implications. But if, as we think, the relationship of thought is very clear between Kelsen and present-day bio-cybernetic jurists, perhaps this is an effect of the same kind as is at play in the transposition to law of the autopoietic paradigm. But before discussing this point, we must now, in a third section, put forward some objections to the autopoietic paradigm, so as to bring out the limits of its descriptive power and its explanatory fertility.

III. Critical Discussion of the Autopoietic Paradigm

A first critical observation that may be formulated in connection with the most radical theories of self-reference concerns the limit of formalisms.

Just as Goedel's proof showed that no mathematical language can manage to be totally looped in upon itself, an essential incompleteness which necessarily presupposes a more enveloping meta-language, likewise the cyberneticist W. R. Ashby, one of the first thinkers of self-organization, managed to prove that absolute autonomy was a logical impossibility (Ashby, 1962; Dupuy, 1982: 226).

Taken literally, reflexive expressions such as "self-programming program" or "self-organizing organization" in fact prove to be contradictory. To be sure, autonomy can be pushed rather far, and systems conceived that have the capacity to change not only the information they process but even their own operating rules. But the capability to change their organizing principles can only be partial: the rules determining these changes are themselves determined, so that the program's control over itself is only partial. Certainly, the program is not immutable, and its rules of change may well be changed in turn, but taking account of these phenomena presupposes a sort of "jump", or qualitative leap to a more complex, more comprehensive level of organization. In other words, the only changes that affect the organization itself — and not only the various states of the system which are nothing but the phases of a constant program — have to be produced from outside the system.

It is so, for instance, with the game of chess with "self-modifying" rules that Hofstadter imagines. It is a particularly subtle game, since the players have agreed that for each configuration of pieces on the chess board there will be a set of rules valid for the next move. Thus, the rules govern the moves, and the moves determine the rules. But as Dupuy, who gives the example, rightly emphasizes, "...this confusion of levels of organization is permissible and conceivable only because there exists a meta-level which itself escapes the recursive loop": this inviolable meta-level is precisely that of the meta-rules that assign a set of rules to each configuration of pieces (Dupuy, 1982: 244). And if one thought of complicating the game still further by associating the meta-rules themselves with configurations of the pieces, one could not thereby escape the limit given by Ashby's theorem. However comprehensive the loop set up may be, one would end up with an unfillable gap, which designates the untouched meta-level: in this case, that of the creative designers who have set up this very special chess game.

Accordingly, the theory of self-organization finds itself facing an essential question: how can self-organized systems claim to change themselves without losing their identity, since they respond totally neither to internal programming, as Ashby has shown, nor to external programming, which would take away any relevance from the very idea of self-organization (Dumouchel and Dupuy, 1983: 97)?

It would seem that with this question the model of self-organization has embarked on a new, decisive path. This path is that of the "relative self-organization" explored by Atlan, "...accepting that the change in

organization may be brought about under the influence, not of a meta-law that would guide the change in constant fashion, foreseen in advance, but under that of random perturbations" (Atlan, 1983: 122). The paradigm of "organizational closure" (Varela) is replaced by that of "organizational chance" or of "order through noise". As Atlan recalls, in a discussion setting him against Maturana: "Changes in organization can come only from outside. There cannot therefore be any self-organization in the strict sense" (Atlan, 1974: 181). But what must be well understood is that, on the one hand, the environment does not act as a central regulator, and on the other, the message received by the system does not necessarily have a destructive effect on it. We must therefore manage to conceive of a form of non-deliberate cooperation by outside chance in the organization of the system, or conversely, a form of autonomy within dependence.

How can chance contribute to creating complexity (the more sophisticated organization of the system, assuring it of increased and active performance) instead of being only a factor of disorganization? To answer this question, which is in no way gratuitous, since it is directed at the process illustrated by the phenomena of evolution of species, one must no doubt again consider the structure of living systems. These are, Atlan explains, located somewhere "between the rigidity of metal and the decomposition of smoke" – hence the title of his magnificent work *Entre le cristal et le fumée* (Atlan, 1979: 5). Living organizations balance two types of property: repetition, regularity and redundancy on the one hand; variety, improbability, complexity on the other (Atlan, 1979: 5). Thus, living automata achieve a form of compromise between determinism and indeterminism, as if a certain quantity of indeterminacy were necessary to the system, beyond a certain degree of complexity, to enable it to adapt to a certain level of noise, a result which Atlan very meaningfully relates to a theorem of von Neumann's games theory (Atlan, 1979: 41).

Accordingly, sufficiently indeterminate systems have a capacity to transform noise, random external perturbations, into an organizing factor which generates increased complexity. In other words, instead of speaking of a "self-programming program", one must analyze the logical nature of systems "where what acts as a program changes unceasingly, in a manner not pre-established, under the influence of random factors of the environment" (Atlan, 1979: 56).

Atlan takes one further step in analyzing such systems by putting forward the idea that if such an effect of "order from noise" is actually possible, it is no doubt because what appears as noise, chance or perturbation to the outside observer is in some way intelligible to or decodable by the system. It is nothing less than the category of "meaning" that is thereby introduced into the analysis. If in fact we agree in this context to define meaning as the effect exercised by a piece of information on the organization of the system, it cannot be denied that information supplied by the environment

has meaning for the system to the extent that it determines a change in its structure, for instance in the form of an adapted response. Accordingly, what appeared to the observer as destruction of information at an elementary level can be seen as creation of information at a comprehensive level. But in this case, the point of view will have been changed, and the system's own key of intelligibility adopted. Nevertheless, since we do not possess this key, we are reduced to speaking of "order through noise", whereas from the system's viewpoint what is happening is translation and processing of intelligible information (Atlan, 1979: 87).

This new conception of self-organization, henceforth understood as "relative" and presupposing only partially self-regulated systems, implies — a point of supreme importance, we feel — a new epistemological viewpoint. Whereas Varela and Maturana's paradigm of "organizational closure" presupposed, as Luhmann also admits, an exclusively internal viewpoint of self-observation, the paradigm of "organizational chance" implies the decentralization of the analysis. Henceforth the point is — and how could it be otherwise — to occupy the viewpoint of the outside observer, an observer who, in his study of systems, takes account of the closure on themselves of systems in respect of their meaning and their purposes (Atlan, 1979: 90). In this way, we feel, one can escape the solipsism of a purely internal viewpoint, without thereby neglecting the self-production of meaning that characterizes the systems studied.

To summarize: the demonstration by such authors as Ashby and Dupuy of the logical limits to the model of self-organization on the one hand, and the development with Atlan, on the other, of a theory of "relative self-organization" pre-supposing systems whose function combines regulation and indeterminacy, certainly facilitate the attempt to apply the self-referential paradigm to social organizations. Before going on in our fourth and last section to that exercise, as far as legal systems are concerned, we must still deal with one author, Castoriadis, who, in thinking about the autonomy of the social on the basis of categories of self-organization, suggests a radical reversal of perspectives.

For Castoriadis, in fact, the autonomous society is the one that frees itself from a heteronomous relationship to its institutions and shows itself capable of rejecting its own system of established meanings. Human societies have for this author always been "autonomous in Varela's sense"; i. e., subject to their institutions and their laws. No doubt this fact was hidden by all sorts of representations that located the origin of the institution in extra-social authorities: the gods, nature, the ancestors, reason or whatever. But the effect of this strategy was to tighten still further the closure of discourse and of institutions, sheltered thereby from all discussion. Such societies are therefore systems of "strong heteronomy" (autonomy in Varela's sense): their informational, cognitive and organizational closure is almost total.

However, on two occasions in the course of history, in ancient Greece and in modern Europe, "autonomous" societies have appeared, that is, ones capable of breaking the closure that characterizes living things, to give themselves laws and institutions which are truly "different". Such societies can be called "self-instituting": by altering their acquired systems of interpretation, they take the risk of looking for norms and forms of organization that no *a priori* programme determines. Castoriadis sees in the emergence of such societies no less than an ontological creation: the appearance of a form which explicitly alters itself as form (Castoriadis, 1983: 421 ff.). This form is, of course, that of democracy, the only mode of government that takes the risk of self-destruction – have we not seen, instead, that autopoietic systems strengthened their "closure" with the intention of escaping entropy?

There, then, is the reversal of perspective: with Castoriadis, autonomy is henceforth openness and no longer closure, the possibility of radical change of program and no longer the complete accomplishment of the program. It is true that Varela and Castoriadis are doubtless not speaking of the same thing: Varela studies life without making pronouncements about social matters, while Castoriadis stresses that autonomy, as he conceives it, appears only with humanity (1983: 438). That does not take away the fact that we are here faced with two totally incompatible conceptions of autonomy: on the one hand, closure of the program, on the other, the continuing invention of meanings.

Ultimately, however, the two theses seem to us equally unacceptable, at least to the extent that they claim to give an account of the actual functioning of human societies. We have already indicated this in various ways in connection with the paradigm of autopoiesis. Regarding the thesis of the self-institution of the social, we shall say that while human societies in fact seem to us to be characterized by their capacity to influence themselves, it is nevertheless the case that in reality this capability is very largely conditioned and reduced by a whole set of constraints. Just as man is at the same time free and determined, societies – even in ancient Greece and modern Europe – are both regulated and regulating. If we really wish to take account of this and not dream of Utopian models, we have to develop concepts that enable us to think simultaneously of constraint, regularity, reproduction *and* freedom, inventiveness, openness. Atlan has supplied some concepts of this type for the analysis of living systems; we think that the *game model* can provide the necessary instruments for analyzing societal systems.

Seventeenth Australasian Tax Teachers' Association Conference

Victoria University of Wellington

27 & 28 January 2005

Legal Autopoiesis and General Anti Avoidance Rules

Geraldine Hikaka[1] and John Prebble[2]

Introduction

The theory of legal autopoiesis is a postmodern attempt to take legal positivism beyond the limitations of earlier theories that ultimately rely on recourse to meta-norms.[3] The theory, being largely the work of German and French scholars, was up to the late 1990s seen by a sceptical Anglo-American academy as[4] "a somewhat obscure and hardly comprehensible doctrine."

This paper discusses autopoietic legal theory principally from the perspective of the model proposed by Niklas Luhmann[5] and applies the theory to the general anti-avoidance rules of the Income Tax Act 1994. Applying abstract theoretical concepts to this specific area of law provides an opportunity to examine the internal coherence of the theory itself.

The paper is in four parts. The first discusses the theory of autopoiesis generally and endeavours to explain its basic concepts. Foundational concepts of the theory posit the legal system, paradoxically, as being "operatively closed" while at the same time "cognitively open". Operative closure is secured by systems' recursively reproducing their own elements self-referentially.[6] Cognitive openness is achieved by "structural coupling", which allows one autopoietic system to "observe" another and to react to it in what Luhmann calls a "systems paradigm".[7]

The second part discusses criticism of the foundational concepts of the theory, particularly the operative closure/cognitive openness paradox that Luhmann asserts. In this

---

[1] LLB Auckland, barrister.
[2] BA, LLB (Hons) Auckland, BCL Oxon, JSD Cornell, barrister, Inner Temple, Professor and former Dean of Law at the Victoria University of Wellington.
[3] For the leading example see Kelsen, Hans, *The Pure Theory of Law*, trans M. Knight 2nd ed (Berkley 1967).
[4] Priban, J and Nelken, D (eds.), *Law's New Boundaries: The consequences of legal Autopoiesis* (Ashgate, Burlington, 2001) 1
[5] Prior to his death in 1998, Niklas Luhmann was Emeritus Professor of Sociology at Bielefeld University.
[6] Baxter, Hugh, "Autopoiesis and the 'Relative Autonomy' of Law" (1998) 19 Cardozo L. Rev. 1987, 1994
[7] Ibid 2037.

connexion, the second part considers Baxter's argument that the concept of an "events paradigm"[8] better explains structural coupling than does Luhmann's systems paradigm.

The third part of this paper acknowledges that tax law's self-referential nature appears to validate Luhmann's assertion that the legal system is operatively closed. Income tax law provides an excellent test bed on which to examine Luhmann's theory because, in the submission of the present authors, income tax law is dislocated and unhinged from its environment in a way other areas of law are not. One of the present authors has termed this disjunction as "ectopia", [9] a word that means "gap" or "dislocation". The fundamental problem of any income tax law is that it cannot tax economic transactions directly. Rather it taxes the legal forms that we use to represent economic transactions.[10] There is a dislocation between the legal form that is taxed and the economic transaction that is the true target of income tax law.

General anti-avoidance rules, however, challenge the coherence of autopoiesis theory as applied to income tax law. In the taxonomy of Luhmann's theory, general anti-avoidance rules are "programmes" that enable structural coupling to occur between, on the one hand, the legal system and on the other hand the political and, economic systems. Luhmann would argue that through the programme of a general anti-avoidance rule, the legal system is able to "observe" the economic system and to react to it by "communications" that are themselves parts of, or possibly new parts of, the legal system itself. On the other hand, if general anti-avoidance rules permit an *exchange* between the systems of law and economy, with the result that the legal system adopts elements that are essentially economic rather than legal, that is, a general anti-avoidance rule acts not merely as a mechanism for observation, then such rules pose a threat to any conclusion that tax law is in fact an autopoietic system.

Part four describes the role of the courts as "paradox managers" (another of Luhmann's terms) showing this role to be critical in maintaining the apparent closed nature of tax law. The authors examine variation in judges' application of general anti-avoidance rules in a number of cases, focusing in particular on *Challenge Corporation Ltd v Commissioner of*

[8] Ibid 2078-2079.
[9] Prebble, John "Income Taxation: a Structure Built on Sand" (2001) Parsons Lecture Sydney University Law School, 14 June 2001. (2002) 24 Sydney Law Review. 301, 306.
[10] Ibid 305.

Inland Revenue[11] and the effect of this variation on legal coding (a further Luhmann expression, explained in part 4).

The paper concludes that although income tax law unmixed with a general anti-avoidance rule is indeed an excellent practical demonstration of Luhmann's thesis, when the legislature adds a general anti-avoidance rule tax law loses what Luhmann would call its "autonomy" and becomes a system that is open rather than operatively closed. The courts manage the extent to which the "substance over form" approach permitted by general anti-avoidance rules allows economic reality to inform legal decisions. Where a decision is informed by economic reality the authors contend that there has been an exchange between the legal system and the economic system, which reveals a porous structure that can hardly be called autopoietic. This conclusion casts doubt on the ability of tax law to maintain its apparent autopoietic nature. Since tax law appears to demonstrate autopoiesis better than all other areas of law, tax law's ultimate failure to qualify as autopoietic calls into question the robustness of Luhmann's theory itself.

The Theory of Legal Autopoiesis

The theory of legal autopoiesis has its genesis in biology, "autopoiesis" translating as "self-making", "self-creation", or "self-production".[12] Concepts borrowed from biology and cybernetics have become established in social systems theory. Through the work of the German sociologists and lawyers, Niklas Luhmann and Gunther Teubner, they colonised legal theory from the 1980s.[13] The appeal of legal autopoiesis is that it offers a more credible solution to the positivist quest for law not to be founded on principles outside of itself than Kelsen's *Grundnorm* theory was able.[14] Social systems theories draw a basic distinction

---

[11] [1986] 2 NZLR 513 (HC, CA, and PC).
[12] Baxter, Hugh, "Autopoiesis and the 'Relative Autonomy' of Law" (1998) 19 Cardozo L. Rev. 1987, 1992.
[13] Ibid 1993.
[14] Teubner, Gunther "Introduction to Autopoietic Law" in Teubner, Gunther, *Autopoietic Law: a New Approach to Law and Society* (1987, de Gruyter, Berlin, New York) 1, 4.

between a system and its environment.[15] Baxter summarises Luhmann's explanation of this distinction and of how it is marked:[16]

> Modern societies are systems of communication that have become differentiated, through a process of social evolution, into various functional subsystems. ... politics, religion, science, education, economy and art, as well as the legal system. The boundaries of these communicative systems are established by a 'code', or system-defining distinction – for example, the legal system's distinction between legal and illegal.

Conventional social systems theory views systems as exchanging inputs and outputs with their environments across system boundaries that are porous. Scholars prefer to see systems as "open".[17] In contrast Luhmann's autopoietic theory describes systems as "operatively closed"[18]meaning that they operate[19]"recursively, producing their own elements self-referentially". This does not result in isolating systems one from another. Luhmann acknowledges[20] that different systems are able to "observe" one another, which paradoxically enables them to be "cognitively open" while retaining their autopoietic nature.

Luhmann asserts[21] that the basic units of a legal system are not legal norms but rather that law is a system of communications and that legal acts are the communicative events that change legal structures. This relationship between legal acts and legal norms is circular,[22] hence autopoietic, and appears to obviate recourse to an external meta-norm. Baxter provides an example:[23]

> When a court communicates a judicial decision, it typically relies on past legal communications, such as prior decisions, and in doing so it creates connective

[15] Baxter, Hugh, "Autopoiesis and the 'Relative Autonomy' of Law" (1998) 19 Cardozo L. Rev. 1987, 1998.
[16] Ibid 1993.
[17] Idem.
[18] Idem.
[19] Ibid 1994.
[20] Idem.
[21] Teubner, Gunther "Introduction to Autopoietic Law" in Teubner, Gunther, *Autopoietic Law: a New Approach to Law and Society* (1987, de Gruyter, Berlin, New York) 1, 3.
[22] Ibid 4.
[23] Baxter, Hugh, "Autopoiesis and the 'Relative Autonomy' of Law" (1998) 19 Cardozo L. Rev. 1987,2006.

> possibilities for future legal communications. In this recursive process, the system reproduces itself as a network of system-specific communications.

Luhmann contends[24] that system boundaries are established internally through the system's own operations and therefore information is system-specific, having meaning only in the terms of the system whose operations produced it. According to Luhmann[25] "coding" takes on the role of characterising a distinction that is specific to law. By "code" Luhmann means[26] that law uses a binary scheme (legal/illegal) in order to structure its own operations and to distinguish them from other facts. A positive value (legal) is applied if a fact conforms to the norms of the system; a negative value (illegal) is applied if a fact violates a norm of the system.[27] Luhmann uses "fact" idiosyncratically, stating:[28]

> A "fact" here is a construction of the system. The system does not acknowledge any external instance that could dictate to it what a fact is, even if the term "fact" can apply to both internal and external phenomena.

The system distinguishes itself from its environment by its own system code, which closes the system off from the other systems in its environment by serving as a "rejection value" for the codes of other systems.[29] Codes only become autopoietically effective as distinctions with the help of a further distinction between coding and programming.[30] According to Luhmann,[31] programming complements coding and fills it with content. Luhmann states:[32]

> Since the values legal and illegal are not in themselves criteria for the decision between legal and illegal, there must be further points of view that indicate whether or not and how the values of the code are to be allocated rightly or wrongly. We shall call these additional semantic elements (in law and in other coded systems) programmes.

---

[24] Ibid 2005-2006.
[25] Luhmann, N, *Law As A Social System*, translated by Ziegert, K, (Oxford University Press, Oxford, 2004), 173.
[26] Ibid 182.
[27] Ibid 183.
[28] Idem.
[29] Baxter, Hugh, "Autopoiesis and the 'Relative Autonomy' of Law" (1998) 19 Cardozo L. Rev. 1987, 2007
[30] Luhmann, N, *Law As A Social System*, translated by Ziegert, K, (Oxford University Press, Oxford, 2004), 193.
[31] Ibid 203.
[32] Ibid 192.

"Legal programmes" include constitutions, statutes, regulations, court decisions and contracts.[33] Programmes must be suitable to direct the allocation of the values "legal" and "illegal". By "illegal" Luhmann does not appear to mean "criminal", nor even "to be condemned". Rather, to put the matter not as Luhmann did but in terms familiar to tax lawyers, he contrasts "having the legal effect that the taxpayer hopes or expects" (legal) with "not having that effect" (illegal). Take, for instance, *Duval & Co Ltd v Federal Commissioner of Taxation.*[34] In that case, in addition to rent, a landlord purported to charge tenants 'contributions to alterations', to help to defray the cost of improvements to certain rented premises. The landlord treated the 'contributions' as receipts of capital. The court held that the contributions were in fact rent, and were therefore assessable as income. From the point of view of the taxpayer in the *Duval* case, the code, "legal", would mean "capital" (and therefore not assessable" and the code "illegal" would mean "rent" or "assessable income". Variation in programmes allows a system to adapt to its environment.[35]:

> A system's coding accomplishes its operative closure and corresponds to the 'self-reference' side of the distinction between self-reference and external reference. A system's programming is the basis for its openness to the environment and its capacity for 'external' reference. A system is thus closed with respect to its code, but 'cognitively' open with respect to its programs. [36]

Luhmann argues that social systems, including the legal system, may be linked through "structural coupling".[37] Essentially, communications occurring in one system do not present as information to another system but rather as "noise".[38] Some of this "noise" may "irritate" the system and "register" in that system's communication.[39] When this happens the "noise" can be processed into information within the system's communication. This may in

[33] Baxter, Hugh, "Autopoiesis and the 'Relative Autonomy' of Law" (1998) 19 Cardozo L. Rev. 1987, 2014.
[34] (1933) 2 ATD 293.
[35] Luhmann, N, *Law As A Social System*, translated by Ziegert, K, (Oxford University Press, Oxford, 2004), 195.
[36] Baxter, Hugh, "Autopoiesis and the 'Relative Autonomy' of Law" (1998) 19 Cardozo L. Rev. 1987, 2011
[37] Ibid 2037.
[38] Idem.
[39] Idem.

turn stimulate the system to transform its structures or categories.[40] Baxter, again, is able to explain Luhmann's proposition succinctly:[41]

> Structural couplings both "admit irritation" and allow systems to "remain indifferent" to the great bulk of each other's communications. In other words, while structural couplings connect communications in the coupled systems, these connections are selective. Even in linking two systems' autopoietic processes, structural couplings presuppose those systems' separation as operationally closed networks of communication. Structural coupling is not fusion.

By way of example Luhmann describes the structural coupling between the two separate systems of law and politics in this manner:[42]

> Legislation is the "place of transformation of politics into law," and conversely, its legally prescribed procedures effect a "legal restriction of politics".

Structural coupling enables mutual and reciprocal "observation" among autopoietic systems,[43] which is seen, for example, when the court in arriving at a legal decision about the meaning of written agreements has observed the economic environment (the commercial context) and has incorporated it into the legal communication itself.[44] However:

> This greater legal openness to the world of economic activity is still selective and still governed by law. ... various court-imposed doctrines – such as unconscionability, duress, undue influence, and other public policy concerns – operate to circumscribe the sphere of contract, as does protective or paternalistic legislation. ...,
>
> ...Even when legal argument incorporates external references to the economic system, it remains an operation of the legal system.[45]

---

[40] Idem.
[41] Ibid 2039.
[42] Ibid 2041.
[43] Ibid 2050.
[44] Ibid 2059.
[45] Ibid 2061, 2062.

Courts hold a central position in Luhmann's theory because they have a legal obligation to render legal decisions,[46] thus placing them in the position of being "paradox managers".[47] Managing decisions involving legal argument is an important means of maintaining an autopoietic legal system:[48]

> Legal argument … is an important form of communication in this centre of the legal system: it presents the courts with alternative paths for decision; it frames the issues for decision by excluding most of the legal environment (and for that matter, most legal communication) as irrelevant, and it connects courts' decisions to past and future legal decisions. Legal argument thus helps reproduce the legal system as a network of communication.

Thus the paradox of the legal system's autopoietic "closed system" and "cognitive openness" can be explained. As Teubner puts it "law can still operate only within its boundaries, but it opens itself to its environment by constructing a legal reality."[49] According to Teubner the legal system's autonomy is only endangered once the "legal coding is in danger of being replaced by criteria of economic utility and political expediency."[50]

Criticism of the Theory of Legal Autopoiesis

Autopoiesis theory suffers from complexity. It is often misunderstood. Vandenberghe criticises[51] it, and particularly Luhmann's version, as being too complex and highly abstracted thereby diminishing its relevance to contemporary debate. On the other hand,

[46] Ibid 2021.
[47] Ibid 2034.
[48] Idem.
[49] Teubner, Gunther "Introduction to Autopoietic Law" in Teubner, Gunther, *Autopoietic Law: a New Approach to Law and Society* (1987, de Gruyter, Berlin, New York) 1, 10.
[50] Ibid 10.
[51] King, Michael, "The Construction and Demolition of the Luhmann Heresy" in Priban & Nelken (eds.) *Laws New Boundaries: The consequences of legal Autopoiesis* above n 1, 123, 124.

King asserts[52] that most critics of the theory demonstrate a lack of understanding and misinterpret the theory, especially the idea of closure.

Baxter takes the view[53] that the two central concepts of "operative closure" and "structural coupling" are an attempt to turn the intuition behind Luhmann's "relative autonomy formula" into a coherent theory. Operative closure, says Baxter,[54] corresponds to "autonomy" and structural coupling to the qualifications expressed in the word "relative". Baxter's connection between autopoiesis theory and the relative autonomy formulation is astute. He notes that this terminology[55] "acknowledges the obvious influences running, in both directions, between law and other social spheres", but does not[56] "by itself characterize, in positive theoretical terms, the relation between law and other social discourses and practices." Baxter describes[57] Luhmann's work on autopoiesis as an attempt to develop the common assumption that law is in some sense relatively autonomous from other social fields into a philosophical theory.

Luhmann maintains that the legal system is operatively closed in the way in which it reproduces and transforms itself by linking up to past legal communications, and by establishing points of connection for future legal communications. He says that "autonomy" is to be understood in this sense.[58] Baxter and others point out that this argument for the legal system's autonomy is tautological:[59]

> Luhmann defines the legal system as operatively closed, and he defines autonomy in terms of operative closure. ... the critics are also correct that Luhmann uses the term "autonomy" idiosyncratically, not with the sense of independence from external influence that the term has traditionally carried.

Luhmann does indeed appear to use the term "autonomy" idiosyncratically. Nevertheless while that usage is not detrimental to its generally accepted meaning of self-

[52] Ibid 124.
[53] Baxter, Hugh, "Autopoiesis and the 'Relative Autonomy' of Law" (1998) 19 Cardozo L. Rev. 1987, 2064.
[54] Idem.
[55] Ibid 1990.
[56] Ibid 1987.
[57] Ibid 2083.
[58] Ibid 2064.
[59] Ibid 2065.

governance, it is clear that Luhmann's theory regards legal self-governance as being conducted entirely self-referentially.

Baxter's critique raises two issues. First, is autopoietic legal theory as posited by Luhmann merely a re-framing of the relative autonomy formula? It appears that the emphasis that Luhmann places on the matter of autonomy as a means to distinguish the legal system from its environment is such that the theory seeks to disengage itself from the formula. Secondly, if this is indeed Luhmann's intention the credibility and consistency of the theory must be examined on its own merits, apart from the idea of relative autonomy.

Two aspects of the theory require further examination: the concept of structural coupling and the interface between coding and programming. The law of contract furnishes examples where observation of other systems through structural coupling goes beyond Luhmann's limits; as far as contract is concerned, the exigencies of commercial reality in the economic system have prompted such significant changes within the legal system that one can hardly call the system "operatively closed" or "autonomous", even in Luhmann's idiosyncratic usage. A thorough analysis of the interface between coding and programming would prudently be the subject of another paper. Nevertheless, the authors venture to examine aspects of this interface through examples in tax cases that raise the question as to whether the distinction between coding and programming can be maintained, particularly where a general anti-avoidance rule is operating, that is, where economic reality rather than law allocates the code value.

Baxter avers[60] that "law and other systems are 'coupled' much more densely and extensively than Luhmann's overview of structural coupling would suggest." He criticises Luhmann's emphasis on "systems" at the expense of "events", viewing Luhmann's exegesis as a departure from a common sense, "event-centered" way of framing the social world.[61]

> Systems and their coded operations are primary for Luhmann; what one might in ordinary conversation describe as an "event" is relevant only so far as it can be seen as the coded operation of an autopoietic system. Thus events that are simultaneously

[60] Baxter, Hugh, "Autopoiesis and the 'Relative Autonomy' of Law" [1987] 19 Cardozo L. Rev. 1987, 2078.
[61] Ibid 2078-2079.

> relevant to two systems of communication–such as a payment (economy) that satisfies a legal judgement (law)–appear in Luhmann's formulation not as a single event, but as two separate operations of two systems.

Luhmann's discussion of structural coupling suggests that connections among systems are few and occur where terms have different meanings in different systems but the difference in meaning allows for reciprocal observation between the two systems that use the term.[62] Baxter disagrees with Luhmann's confining of coupling between law and politics to constitutions, legislation and enforcement of legal judgments and confining coupling between law and economy to property and contract. Baxter states:[63]

> Similar analysis would establish a whole series of other "structural couplings" between law and economy. Obvious examples would be the idea of the corporation, various notions of tax law, "competition" in antitrust law, and … "liability"–all of which are significant in both legal and economic discourse, and all of which allow reciprocal "observation" between the two systems. One could discover innumerable other couplings between law and economy simply by following the path of the law-and-economics movement.

There is much force in Baxter's view that Luhmann's emphasis on systems has resulted in too narrow a view of the amount of structural coupling that occurs between systems, particularly between law and the economy. An "events" paradigm certainly has greater capacity to discern and then to describe opportunities for coupling. It would seem that Luhmann's focus on systems and on the outcome of the effect of coupling on those systems has misinformed his view that connections between systems are few.

Minimising the amount of structural coupling occurring between systems appears critical to maintaining the integrity of the theory. If more and deeper coupling occurs than the theory acknowledges, thereby increasing the risk of exchange, and not merely an observation, autopoietic theory is in danger of collapsing into a conventional social systems theory view of systems that are porous and hence "open".

[62] Ibid 2076.
[63] Ibid 2077.

The present authors agree with Baxter and would go further. They are sceptical as to whether autopoiesis theory has much to offer as a perspective on most of the legal system. For example, consider, from the law of contract, two developments in the law of consideration, one ancient and one modern.

Until the seventeenth century the common law would not enforce executory contracts. One could sue on deeds,[64] or for debt,[65] or for misfeasance,[66] but a simple bilateral exchange of mutual promises, each promise to act or to pay in the future, was unenforceable. As Smith has explained,[67] the reason is that pre-modern societies thought in terms of the factual rather than the abstract. People thought in terms of categories of which the properties were experienced empirically, not of concepts of which the properties were future-oriented and logically derived. Moreover, in England, the forms of action at common law constituted a straitjacket that preserved laws that reflected these old forms of thinking well into the Enlightenment.[68]

The pressures of increasingly sophisticated commerce meant that there had to be a change. The watershed was a relatively simple dispute: Morley promised to buy all the wheat grown in Slade's field, at a price that the parties agreed. When the wheat was harvested Morley dishonoured his promise. As the law then stood, Slade had no remedy, but he issued a writ anyway. There was a special sitting of all the judges of the courts of Kings Bench and Common Pleas. To put their decision in modern terms, they held that a promise of future action or payment was sufficient consideration to support a binding contract.[69] Thus, *Slade's Case* has a strong claim to being the foundation of modern commercial law.

The second example is *Antons Trawling Co Ltd v Smith* (2003),[70] which involved a variation of contract. The dispute played out against New Zealand's licensing system for

[64] Burrows, JK, Finn & Todd *The Law of Contract in New Zealand* (Wellington 2d ed 2002) 3.
[65] Kiralfy, AKR, *Potter's Historical Introduction to English Law and its Institutions* (London, 4th ed 1962) 452 – 454.
[66] *Bukton v. Tounesende (The Humber Ferry Case)*(1348), translated in Baker & Milsom's *Sources of English Legal History* (Butterworth's, London, 1986) 358.
[67] Smith, JC, "The Unique Nature of the Concepts of Western Law" (1968) 46 Canadian Bar Review 202 – 207.
[68] See generally, Kiralfy, AKR, *Potter's Historical Introduction to English Law and its Institutions* (London, 4th ed 1962) Part Three, Chapters 2 and 6.
[69] *Slade's Case*, (1602) 4 Co Rep 91a, Yelv 21, Moore KB 433, 667.
[70] [2003] 2 NZLR 23, CA.

commercial fishing. The system operated by allocating transferable quotas of fish that quota-owners were entitled to take. One way of obtaining more quota was to discover new fishing grounds. Smith was the captain of a fishing trawler belonging to Antons Trawling Co Ltd. Antons promised Smith that he would own ten per cent of any quota that he might obtain for Antons by virtue of discovering new grounds. One of the strands of the case proceeded on the basis that searching for new fishing grounds was already among Smith's duties pursuant to his contract of employment. That is, there was no consideration for Antons' promise to transfer quota to Smith, since Antons already paid Smith an agreed salary for his work as a captain.[71]

The Court of Appeal held that the promise was nevertheless binding, holding that in the context of modern commercial transactions consideration is not necessary for a variation of contract to be binding, even where the variation does not require the promisee to perform additional duties.[72]

In terms of Luhmann's analysis, the rules about consideration are part of the content of the code of an autonomous legal system. In *Slade's Case* and in *Antons Trawling Co Ltd v Smith* commercial pressures and realities, being part of the economic system, were "noise" that "irritated" the legal system and that caused it to change its point of view in applying the relevant value of its code. The present authors are not convinced. What does Luhmann's terminology add to saying that in both cases the courts recognized that the norms of the legal system did not reflect commercial reality (or even common sense) and adjusted the rules to reflect that reality (or common sense)?

Law and the economy are qualitatively different. Law is a system of norms, while the economy is composed of facts. But it is both accurate and readily understandable to say that the judges in the two cases were open to and were influenced by the facts of the economic system. Luhmann does not appear to say anything different, though he puts it in obscure language.

---

[71] *Harris v Watson* (1791) Peake 102 Lord Kenyon, *Stilk v Myrick* (1809) 2 Camp 317 Lord Ellenborough, *Cook Islands Shipping Co Ltd v Colson Builders Ltd* [1975] 1 NZLR 422 Mahon J.

[72] The court stipulated certain conditions for such a variation to be binding. Also, it alternatively based its decision on a second *ratio decidendi* derived from *Williams v Roffey Bros & Nicholls (Contractors) Ltd* [1991] 1 QB 1 and *Attorney-General for England and Wales v R* [2002] 2 NZLR 91.

It is true that both cases might have come to conclusions that were opposite to the judgments that were in fact delivered. The courts might have held that contrary precedents prevented the developments for which counsel for Slade and Smith argued. Such holdings might have supported an argument that the legal system is "operatively closed", and that changes are made at the interface of law, politics, and the economy in the form of legislation, as has occurred in other areas of the law of contract since the seventeenth century.[73] But to what end might one make this argument? It establishes no more than the obvious point that while the courts consider and are influenced by their economic and social context, (the economic and social systems in Luhmann's terms) their function is to preside over a system of norms, norms that the judges have limited scope to change. Radical changes are for Parliament.

Does this analysis of *Slade's Case* and *Antons Trawling Co Ltd v Smith* vindicate the authors' argument that Luhmann's perspective seems to add little to a conventional view of the law? The cases do, after all, consider one part only of consideration, a single aspect of the law of contract. Are the present authors fair to generalize from such a small sample?

The problem is, of course, that to refute Luhmann in a statistically robust manner one would have to address a huge swathe of the canon: as task that is impractical. The authors also plead that while scholars of autopoiesis are long on description they are remarkably thrifty when it comes to the exposition of particular examples. For instance, of the authors in Teubner's *Autopoietic Law: a New Approach to Law and Society*,[74] only Thomas Heller[75] essays a consideration of a specific field of law in any depth, namely the law as to bilingual and bicultural education.

Heller embarks on his theme with modest expectations. He explains that: "It is problematic whether the legal system has the capacity to reproduce itself as an autonomous

[73] Eg, from the United Kingdom, The Statute of Frauds 1677, Lord Tenterden's Act 1828, the Unfair Contract Terms Act 1977; and from New Zealand, the Sale of Goods Act 1908, the Illegal Contracts Act 1970 and the Contracts (Privity) Act 1982.
[74] Teubner, Gunther, ed, *Autopoietic Law: a New Approach to Law and Society* (1987, de Gruyter, Berlin, New York).
[75] Heller, Thomas "Accounting for Law", ibid 283.

system in the face of a reconfiguration of concrete practices in both the economic and cultural subsystems discordant with existing legal representations."[76] Heller's writing shows that his doubts were well founded. The essay discusses conflicting economic, political, and cultural forces and their effect on the law relating to bilingual and bicultural education, but it does not apply itself at all to analysing lawyers' law, such as examples of judicial reasoning or even the text of a statute. It concludes with a number of high-level stipulations, such as, "The possibility of a qualitative shift in the underlying basis of social relations calls into question the dominant images of liberal capitalism, social democracy, and an international order founded on the interaction of national states."[77] The reader concludes that Heller's attempt to clothe his discussion in autopoietic discourse obscures rather than illuminates the points that he makes, cogent though they are.

The present authors do not wish to be unfair to Heller in singling out his essay in this manner. At least, unlike his fellow authors in *Autopoietic Law: a New Approach to Law and Society,* Heller descends from the generalizations that mark most writing on autopoiesis and attempts to see whether the theory actually works when applied to a specific area of law. On the other side of the argument, the present authors submit that it is not hard to cite familiar examples of legal reasoning that, like *Slade's Case* and *Antons Trawling Co Ltd v Smith,* demonstrate that most of law is far from a closed, autopoietic system, without needing to subject these examples to full analysis. To name three, take the attractive nuisance fiction from torts,[78] the doctrine of corporate personality from company law,[79] and the defence of automatism from criminal law.[80]

Tax Law and Legal Autopoiesis Theory

Curiously, if there is one area of law that can offer a lifeline to Luhmann it is an area that he neglects, namely income tax law. Above all other law, income tax exhibits the qualities of operative closure, recursiveness, and self-reproduction that Luhmann and Teubner see in the legal system in general. One of the present authors has set out the reasons and the evidence

[76] Ibid 301.
[77] Ibid 310.
[78] *Sioux City & Pacific Rly Co v Stout* (1873) 17 Wall. 657 (US SC), *City of Pekin v McMahon* 154 Ill 141 (1895), and *United Zinc v Bruitt* 258 US 268, 275 (1921) per Holmes J. But see *Addie v Dumbreck* [1929] AC 358 HL.
[79] *Salomon v Salomon & Co Ltd* [1897] AC 22 HL, *Aspro Ltd v Commissioner of Taxes* [1930] NZLR 935 PC.
[80] *Bratty v Attorney-General (Northern Ireland)* [1963] AC 386.

in a number of articles.[81] In short, income tax law is by its nature dislocated from its subject matter in a manner that distinguishes it from the law in other areas. There is a gap, or ectopia, between tax law and the facts to which it relates.[82] This ectopia comes about for a number of reasons. The primary one is that tax law cannot tax the economic facts of transactions and investments directly, but must instead address a legally defined surrogate, which is a legal simulacrum of those facts. The very definition of "income", which is fundamental to income tax law, is a simulacrum rather than a direct definition of the economic phenomenon that is the target of tax law. As the author has put it, a "true, economic, business profit, which would be [a] proper subject of the tax base if we could ever get at it, is removed from its legal simulacrum by an ectopia".[83]

A second reason for the ectopia of income tax law is that it is based on a number of fictions. They are of two types: counter-factual fictions and counter-legal fictions. The former are fundamental to income tax law and are ineliminable. The primary example is the assumption that there is a logical, real-world distinction between capital and revenue.[84] A second is that all income can be attributed to a physical geographical source.[85]

Counter-legal fictions are creatures of statute, enacted to bring income tax law closer to the economic facts to which it relates, or to relieve inequities that arise from the application of crude principles of unrefined income tax law, or to close loopholes. Respectively, examples of the three types include: rules to recharacterize hire purchase

[81] Prebble, John, "Ectopia, formalism, and anti-avoidance rules in income tax law" (1994) in W. Krawietz N. MacCormick & G.H. von Wright (eds) *Prescriptive Formality and Normative Rationality in Modern Legal Systems, Festschrift for Robert S. Summers*, Duncker and Humblot, Berlin, 367-383; "Philosophical and design problems that arise from the ectopic nature of income tax law and their impact on the taxation of international trade and investment", (1995) 13 *Chinese Yearbook of International Law and Affairs*, 111-139, reprinted as "Ectopia, tax law, and international taxation" [1997] *British Tax Review* 383; "Can income tax law be simplified?" (1996) 2 *NZ Journal of Taxation Law and Policy* 187; "Should tax legislation be written from a principles and purpose point of view of a precise and detailed point of view?" [1998] *British Tax Review* 112; "Income Taxation: a Structure Built on Sand" (2002) 24 *Sydney Law Review* 301; see also "Why is tax law incomprehensible?" (1994) *British Tax Review* 380-393.

[82] Prebble, John, "Ectopia, formalism, and anti-avoidance rules in income tax law" (1994) in W. Krawietz N. MacCormick & G.H. von Wright (eds) *Prescriptive Formality and Normative Rationality in Modern Legal Systems, Festschrift for Robert S. Summers*, Duncker and Humblot, Berlin, 367-383

[83] Prebble, John "Income Taxation: a Structure Built on Sand" (2002) 24 *Sydney Law Review* 301, 306.

[84] Ibid 309.

[85] Prebble, John "Philosophical and design problems that arise from the ectopic nature of income tax law and their impact on the taxation of international trade and investment", (1995) 13 *Chinese Yearbook of International Law and Affairs*, 111-139, reprinted as "Ectopia, tax law, and international taxation" [1997] *British Tax Review* 383.

transactions so that they are taxed as credit sales, which are their economic equivalents, rather than as leases;[86] close company rules that qualify small companies of few members to be taxed as partnerships;[87] and controlled foreign company regimes, which attribute the income of foreign companies to their owners who are resident in the taxing jurisdiction.[88] Counter-legal fictions are not ineliminable in the same way as tax law's counter-factual fictions. That is, tax law could exist without them. Nevertheless, these rules that recharacterize transactions or that in some other way treat transactions or structures otherwise than according to their legal form can be abandoned only at the cost of an increase in ectopia, that is, an increase in the distance between tax law and its subject matter.

Counter-factual fictions and counter-legal fictions exemplify the pronounced autopoiesis of income tax law in different ways. Counter-factual fictions are examples of the autonomy and operative closure of law. Tax law is so independent of its economic context that some of its most basic rules run counter to economic fact. Secondly, counter-factual fictions trigger recursive, self-referential judicial reasoning, which is best seen in cases on the distinction between capital and revenue. For instance, Lord Wilberforce felt obliged in *Tucker v Granada Motorway Services Ltd*[89] to characterize a certain lease that was central to the court's reasoning as a capital asset because that is the way that tax law treats leases, notwithstanding that the lease in question was of negative value, could not appear in the taxpayer's balance sheet, and was non-assignable.[90]

Statutory counter-legal fictions are also examples of recursive self-reproduction. Parliament creates notional structures and relationships that, as Richardson J put it, "have no reality under the statute except in relation to income tax"[91] and then requires the fisc to calculate income and tax on the basis of those fictional structures and relationships. In terms of Isenbergh's celebrated *mot*, in much of tax law"[S]ubstance *is* form and little else; there is no natural law of reverse triangular mergers."[92]

[86] Eg Income Tax Act 2004 s FC 10.
[87] Ibid subpart HG.
[88] Ibid subpart CG.
[89] [1979] 2 All ER 801 HL /T.C Leaflet No 2727
[90] Ibid 813. For a longer discussion and other examples, see Prebble, John "Income Taxation: a Structure Built on Sand" (2002) 24 *Sydney Law Review* 301, 309-311.
[91] *Challenge Corporation Ltd v Commissioner of Inland Revenue* [1986] 1 NZLR 513, 552 line 56, CA.
[92] Isenbergh, Joseph "Musings on Form and Substance in Taxation" (1982) 49 Chicago Law Review 859, 879.

The reason for this peculiarity of tax law is its ectopia. Because tax law only imperfectly reflects and captures the transactions and structures that are its target, taxpayers find it relatively easy to restructure their affairs, to the end that tax law categorizes and treats them differently. The counter-legal fictions just described are tax law's response. An intriguing question is, "How do general anti-avoidance rules fit into this analysis?"

General Anti-Avoidance Rules

From the perspective of Luhmann's theory, a general anti-avoidance rule appears in some senses to be the epitome of autopoiesis. This is particularly so in respect of general anti-avoidance rules that empower the Commissioner to reconstruct the taxpayer's transactions[93] or to shift fiscal benefits or tax liability from one taxpayer to another.[94] Such provisions empower the Commissioner to attribute transactions to taxpayers that did not occur as far as the rest of the law is concerned, and then to tax those transactions instead of the transactions that actually occurred. It is hard to think of a more dramatic example of reasoning that is recursive and self-reproductive.

At the same time, however, the operation of a general anti-avoidance rule makes the tax regime not just "cognitively open" (in Luhmann's sense of taking notice of what is going on in adjacent social systems) but "operatively open", rather than "operatively closed", which Luhmann claims to be characteristic of law. The tax regime becomes operatively open to the economic system in that the essence of a general anti-avoidance rule is that it reverses tax law's ordinary emphasis on form and requires the Commissioner and the courts to look to the economic substance of the transaction in question.[95] In terms of autopoietic theory, the general anti-avoidance rule may be viewed as a change in legal programme that facilitates Luhmann's "structural coupling" between the legal system and the economic system.[96]

---

[93] Eg Income Tax Act 2004 s GB 1.
[94] Eg Income Tax Assessment Act (Aust) 1936 s 177F.
[95] Eg *Hadlee and Sydney Bridge Nominees v CIR* [1991] 3 NZLR 517, 524 per Cooke P (1991) 13 NZTC 8,116 (CA), affirmed on another point [1993] 2 NZLR 385 PC.
[96] Ost, Francois "Between Order and Disorder: The Game of Law", in Teubner, Gunther, *Autopoietic Law: a New Approach to Law and Society* (1987, de Gruyter, Berlin, New York) 70, 75.

One might go further. The authors submit that that the operation of a general anti avoidance rule involves more than a mere "structural coupling" between the legal and the economic systems. Rather, general anti-avoidance rules have the objective of making the legal system truly reflect the economic system as regards the tax effect of the transactions to which they apply. In its Report to the Treasurer and Minister of Revenue the Committee of Experts on Tax Compliance explained:[97]

> Tax incentives, concessions and loopholes in tax legislation create intended and unintended tax consequences for certain types of economic activity. ... Ideally, anti-avoidance rules should act as a deterrent to tax avoidance arrangements.

Tax, as an "event", provides the opportunity for at least three systems to observe one another, these being the law, the economy and the political system. Each system has its own perspective on tax as an event, but there is a question as to whether and to what extent each system's view of the "fact" of tax is dictated by a "fact" in another system through the process of structural coupling. The critical issue is whether a system's programme has been so informed in its point of view by another system that it is the other system that in reality allocates the values of the code.

Income tax law's distinction between capital and revenue severely limits the influence of the economic system when there is structural coupling between law and economy. The distinction distorts economic reality and emphasises the autopoietic character of much of tax law. In contrast, general anti-avoidance rules act as a kind of gateway that makes tax law (which is otherwise autonomous) vulnerable to the influence of other systems, in particular to the economic system. General anti-avoidance rules therefore enable the economic system to become structurally coupled with tax law around the event of tax. Indeed, it is a central thesis of this article that a general anti-avoidance rule is apt to make the conjunction of the legal and economic systems a more intense occasion than Luhmann contemplates by a structural coupling where one system merely "observes" another.

[97] *Report of the Committee of Experts on Tax Compliance* Inland Revenue Department Tax Policy Division Chapter 6: "Tax Mitigation, Avoidance and Evasion" (1998) 6.29. http://www.taxpolicy.ird.govt.nz/publications/files/html/coe/chapter6_1.htm (at 31 March 2004).

The uncertain ambit of general anti-avoidance rules enhances this intensity. To quote the Committee of Experts again:[98]

> If a general anti-avoidance provision is to be effective, it cannot be precise. Although this feature of an anti-avoidance provision means less certainty for taxpayers, the committee believes that this cost is outweighed by the benefit provided by the flexible wording of the general anti-avoidance rule, allowing the court to address new and different types of avoidance arrangements. Again this helps to preserve the robustness of the tax system.

Luhmann would say that general anti-avoidance rules facilitate his conception of "cognitive openness" of the legal system by virtue of their imprecision. They open tax law to the economic facts of different types of avoidance arrangements and results in strengthening that part of the legal system that comprises tax law. The process gives the legal system a mechanism whereby it can code a tax event by its substance rather than by its form. The open-ended nature of the legislation makes the courts more effective in their role of "paradox managers" in allowing them to address new and different types of avoidance arrangements by relating those arrangements to their underlying economic substance.

Francois Ost[99]explains what it is that autopoietic theory views as being at stake in this paradoxical "balancing act":[100]

> An autopoietic system, defined as functionally (normatively in the case of law) closed, is also a system open to the environment, a source of information for it. The normative closure is thus parallelled by cognitive openness. A legal system can be programmed in such a way as to render itself deliberately dependent on the evolution of outside circumstances; a change in the program itself under pressure from the environment is likewise possible, as long as the system does not in this search for optimum adaptation lose control over its transformations.

[98] Ibid 6.34.
[99] Francois Ost "Between Order and Disorder: The Game of Law", in Teubner, Gunther, *Autopoietic Law: a New Approach to Law and Society* (1987, de Gruyter, Berlin, New York) 70, 75.
[100] Ibid 75.

This is a critical point. It would appear that general anti-avoidance rules are a mechanism by which the legal system has programmed itself so as to be deliberately dependent on circumstances within the economic system that could give rise to tax avoidance. The legal system's binary code value of "legal/illegal" will be applied to a tax event. The extent to which the legal system deliberately depends on the economic system in deciding whether a tax event is avoidance (and thereby illegal)[101] is critical as to whether tax law within the legal system remains genuinely autopoietic. If the code value is allocated from the point of view of a "fact" as decided by the economic system it is arguable that there has been an exchange between the two systems, not merely an observation. As Francois Ost states:[102]

> In combining closure and openness in this way, the legal system cannot lay claim to complete self-determination.

It appears to be a feature of general anti-avoidance rules that they cause legal code to be allocated by criteria of economic substance. Teubner sees developments of this kind as threats to the legal system[103], though paradoxically, the reverse may be true, at least in respect of the general anti-avoidance rule and its relationship to tax law. To test whether and how completely autopoietic and self-determining tax law is within the legal system it is necessary to examine how the courts as "paradox managers" have applied the general anti-avoidance rules.

Courts as "Paradox Managers"

The foundation case in the area of tax avoidance is *Duke of Westminster v IRC*.[104] The Duke sought to pay his gardener by way of an annuity paid weekly. The Duke's objective was to divert income to the gardener before it reached his Grace's hands. This strategy would mean

[101] Within tax law, Luhmann's binary code of legal/illegal cannot refer to "guiltless" or "criminal" alone, but are compendious terms that include, for instance, the distinction between a transaction that takes effect according to its terms and a transaction that is void pursuant to an anti-avoidance rule.
[102] Ibid 75.
[103] Teubner, Gunther "Introduction to Autopoietic Law" in Teubner, Gunther, *Autopoietic Law: a New Approach to Law and Society* (1987, de Gruyter, Berlin, New York) 70, 75
[104] [1936] AC 1 HL.

that rather than the Duke deriving income and suffering surtax on it before paying wages out of the income the money would go straight to the gardener. The Commissioner argued that in substance the annuitant was serving the Duke for the equivalent of his former salary or wages and therefore the annuity should be treated as salary or wages.[105] The court, however, decided the case according to the legal form of the transaction. Lord Tomlin held:[106]

> Every man is entitled if he can to order his affairs so as that the tax attaching under the appropriate Acts is less than it otherwise would be. If he succeeds in ordering them so as to secure this result, then, however unappreciative the Commissioners of Inland Revenue or his fellow taxpayers may be of his ingenuity, he cannot be compelled to pay an increased tax. This so-called doctrine of "the substance" seems to me to be nothing more than an attempt to make a man pay notwithstanding that he has so ordered his affairs that the amount of tax sought from him is not legally claimable.

Clearly the court was not prepared to scrutinize the transaction from an economic perspective. It was only prepared to look at legal form and not the economic reality of the payment. Lord Tomlin's judgment demonstrates a distinctly autopoietic response, entirely self-referential to legal norms, with a refusal to allocate the code value by any reference other than to law's own existing categories. This approach is typical of common law jurisdictions. As Tipping J put it in *A Taxpayer v C of IR*:[107]

> In New Zealand, liability for income tax should be determined according to law, not according to elastic notions of economic reality; albeit that in cases of sham and avoidance a different approach may be required.

Tipping J's words assert the legal system's claim, almost its right or destiny, to determine liability for income tax. According to Luhmann's theory, this function would be found in the coupling of law and politics that has produced the applicable income tax legislation. Nevertheless the statements of both Lord Tomlin and Tipping J indicate resistance to coupling with the economic system, even in circumstances where sham and

[105] Ibid 19.
[106] Idem.
[107] (1997) 18 NZTC 13,1350, 13,360.

avoidance are present. Admittedly, Tipping J acknowledges that a different approach may be required for sham and avoidance, but it seems implicit in his statement that, in spite of "observation" of the economic system by coupling, the decision will be made by reference to the legal system and its categories rather than by an approach that accommodates economic reality.

Baxter's assertion that an events paradigm is a more accurate exposition of autopoiesis theory would mean that tax liability provides numerous opportunities for structural coupling between the legal and economic systems. The degree to which judicial reasoning acknowledges economic reality may go some way to exposing how porous, or otherwise, the structural coupling between these two systems really is. It may be that if the coupling is porous, rather than merely observational, it could give a different meaning to Tipping J's statement that tax liability is determined [entirely] according to law.

In Luhmann's theory general anti-avoidance rules are a programme initiated by the legal system the better to observe the economic system. Programmes compliment coding and fill coding with content[108] but it appears that general anti-avoidance rules have the effect of making the autopoietic nature of the legal system vulnerable to substantive influence from the economic system. It becomes critical to sustaining an argument in support of Luhmann's autopoietic theory that the courts, in deciding whether there is tax avoidance, should refer recursively to tax law in constructing a "fact". To maintain a truly autopoietic nature the legal system must resist the economic reality of the transaction from determining which of the binary code values of legal/illegal is accorded to the transaction.

The United Kingdom has never enacted a general anti-avoidance rule. As a result, Lord Tomlin's judgment in the *Duke of Westminster* case employed the autopoietic reasoning that is familiar in tax law. New Zealand has long had a general anti-avoidance rule, but it was not relevant in the case of *A Taxpayer*. Accordingly, Tipping J similarly hewed to an autopoietic line. Cases where a general anti-avoidance rule is in play present a very different picture. New Zealand's rule has undergone many changes over the years, but for present purposes its present form well illustrates a fairly typical form of the rule. Section BG 1 of the

[108] Luhmann, N, *Law As A Social System*, translated by Ziegert, K, (Oxford University Press, Oxford, 2004), 203

Income Tax Act 2004 and its companion definition of "tax avoidance arrangement" in section OB 1 read:

> **BG 1 Tax avoidance**
> *Avoidance arrangement void*
> (1) A tax avoidance arrangement is void as against the Commissioner for income tax purposes.
> *Reconstruction*
> (2) Under Part G (Avoidance and Non-Market Transactions), the Commissioner may counteract a tax advantage obtained by a person from or under a tax avoidance arrangement.
>
> **Tax avoidance arrangement** means an arrangement, whether entered into by the person affected by the arrangement or by another person, that directly or indirectly-
> (a) Has tax avoidance as its purpose or effect; or
> (b) Has tax avoidance as 1 of its purposes or effects, whether or not any other purpose or effect is referrable to ordinary business or family dealings, if the purpose or effect is not merely incidental:

Cases where the court has held that a general anti-avoidance rule governs may be seen either as examples of erosion of the autonomy of the legal system or as cases where economic structures and categories have been incorporated into legal reasoning so as to inform the legal coding that comprises its building blocks. One example is *Mangin v Commissioner of Inland Revenue,*[109] where the court ignored legal code of the trust structure in favour of economic reality.

Challenge Corporation Ltd v Commissioner of Inland Revenue[110] raised a question about the relationship of the general anti-avoidance rule (then section 99, Income Tax Act 1976) to a specific concessionary provision, namely section 191 of the Act. Section 191 allowed companies within a group to set losses and profits off within the group, being taxed on the net sum that was left. To qualify, companies had to be within the same group at the end of the tax year. There was a specific anti-avoidance provision aimed at cases where a

---

[109] Eg *Mangin v Commissioner of Inland Revenue* [1971] NZLR 591
[110] *Challenge Corporation v Commissioner of Inland Revenue* [1986] 2 NZLR 513

company entered and left a group after a short time, meanwhile taking advantage of section 191.

The taxpayer purchased all the share capital in a company named Perth, thereby taking Perth into the Challenge group, where it remained at year-end. Perth had no assets or debts but a large loss, which Challenge set off against its own profit.[111] In effect Challenge purchased tax losses. The Commissioner disallowed the deduction on the basis that it arose by reason of an arrangement purporting to alter the incidence of tax and was, therefore, voided vis-à-vis the Commissioner by section 99.[112]

In the High Court the taxpayer argued that section 99 could not apply in circumstances where detailed provisions in the statute itself covered the circumstances that obtained (that is, the sharing of tax losses within a group) if (a) the taxpayer had complied to the letter with such provisions and (b) the provisions included their own anti-avoidance rules.[113] Barker J held:[114]

> In my view, whilst s 99 applies prima facie to the transaction, the restriction on its operation, developed by cases concerned with its predecessor or similar provisions, prevents its application to a case such as the present, where the taxpayer has observed to the letter an express and complicated provision in the Act, even where the taxpayer's aim was to reduce its tax liability.

In deciding which provision took precedence, Barker J found it necessary first to determine whether the transaction was an "arrangement". To do so his inquiry went beyond an examination of the statutory provisions. His opening statement indicates that his Honour considered commercial reality, that is, the facts of the economic system. He said:[115] "The transaction under consideration was an 'arrangement' which would not normally have been entered into by the objector."

[111] Ibid 513.
[112] Idem.
[113] Ibid 519.
[114] Ibid 529, Line 36.
[115] Ibid, 525 line 22.

His Honour went on to list the commercial steps that were taken to effect the transaction, including contracts for the sale and transfer of shares. The key question here is: to what system should the court refer the actions of the taxpayer when it is deciding whether those actions were normal or extraordinary–the economic system or the legal system? A further statement indicates that the norms to which His Honour referred were from the world of commerce:[116] "It is difficult to see why the objector would wish to purchase the shares in a bankrupt company other than to take advantage of its tax losses."

The court's decision as to whether the transaction was an "arrangement" involved more than a mere "observation" of the economic system. Economic norms were imported into the legal system to inform the decision making process. This is evidenced by the transaction being viewed as "not normal" in the usual commercial activities of Challenge. Luhmann's assertion[117] that one system does not acknowledge any external element that could dictate to it how it constructs a "fact" rings hollow here. The transaction is only an "arrangement" (and therefore a "fact" that violates the norms of the legal system) because the commercial transaction was an extraordinary transaction for Challenge according to commercial norms. The legal form of the transaction complied with the letter of the statute and therefore complied with the norms of the legal system and could expect to be allocated the code value "legal". Nonetheless his Honour considered the nub of the argument lay elsewhere, stating:[118]

> In my view, the transaction prima facie comes within the purview of s 99; that is not the end of the matter. I must now consider whether the fact that the objector complied to the letter with specific provisions in the statute relating to grouping and carrying forward of tax losses negates the prima facie effect of s 99.

By this recursive reasoning, Barker J was re-entered the autopoietic polity of tax law, confining the legal system's coupling with the economic system to a mere observation of

[116] Ibid 525 line 44.
[117] Luhmann, N, *Law As A Social System*, translated by Ziegert, K, (Oxford University Press, Oxford, 2004), 183.
[118] *Challenge Corporation v Commissioner of Inland Revenue* [1986] 2 NZLR 513, 525, lines 50-52.

commercial norms. The binary code value was ultimately applied according to the norms of the legal system, consideration of avoidance being limited to that already present in the anti-avoidance provision in the grouping and loss carrying forward parts of the statute.[119]

In the Court of Appeal Richardson J's judgment indicates an adept understanding of the role of courts as "paradox managers." He demonstrates tax law's autopoietic capacity to subsume external observations to tax law concepts, thereby maintaining the legal system's closure by self-recursive reasoning.

In autopoietic terms Richardson J begins by acknowledging evidence of historical structural coupling between the legal system and economic and political systems. He states:[120]

> The upsurge of tax planning dates from the 1950s and was a product of higher tax rates which with increasing affluence and inflation were felt by more taxpayers as ever increasing reliance was placed on income tax both as a generator of Government revenue and as a means of serving wider and at the same time specific economic and social policies.
>
> ....
>
> Tax is a highly significant factor in business decision making and family property planning and specific anti-avoidance provisions necessarily have limited impact. Against that background it is not surprising that the legislature should raise a general yardstick by which the line between legitimate tax planning and improper tax avoidance is to be drawn.

This statement indicates that activities in the legal system's environment around the event of tax, are, in the avoidance scenario, more than "noise" and have become an "irritation" to the legal system. Richardson J is clear as to what the response of the legal system should be, asserting the "paradox management" role of the court. He states:[121]

> Commercial decisions are inevitably influenced by taxation implications and

[119] Ibid 529, lines 29-31.
[120] Ibid 545, lines 23, 48.
[121] Ibid 548, lines 38, 549, line 22.

> ... in a vast range of ordinary business activities conducted within the framework of the income tax legislation the minimising of tax and the saving of tax could never be regarded as a merely incidental purpose or effect. In these circumstances it becomes particularly important to determine the relation between s 99 and the other specific provisions in the legislation under which tax changes may occur. The legislation having failed to provide specific tests in this regard, it falls to the Courts to do so as a matter of statutory construction.
>
> ...
>
> In the end the legal answer must turn on an overall assessment of the respective roles of the particular provision and s 99 under the statute and of the relation between them.

Richardson J's reasoning clearly seeks to distinguish the legal system from its environment. In acknowledging factors outside the legal system that interface with the tax event he subsumes their ability to influence the allocation of the values of the legal code by confining the ambit of the enquiry to statutory construction, the statute being a legal programme. In this way the legal system, as Baxter has observed,[122] reproduces itself as a network of system-specific communications, though this analytical framework momentarily appears to be under threat when later in the judgement Richardson J considers the scheme and purpose of the legislation. Richardson J states:[123]

> It is obviously fallacious to assume that revenue legislation has a totally coherent scheme, that it follows a completely consistent pattern, and that all its objectives are readily discernible. There is force in the thesis that in many respects the tax base is so inconsistent and contains so many structural inequities that a single general anti-avoidance provision such as s 99 cannot be expected to provide an effective measure by which to weigh the exercise of tax preferences... Nevertheless, that emphasis on trying to discern the scheme and purpose of the legislation is likely to provide the legal answer to the relation between s 99 and other provisions of the Act that best reflects the intention of Parliament as expressed in the statute.

[122] Baxter, Hugh, "Autopoiesis and the 'Relative Autonomy' of Law" (1998) 19 Cardozo L. Rev. 1987, 2006.
[123] *Challenge Corporation v Commissioner of Inland Revenue* [1986] 2 NZLR 513, 549, line 34.

In the end, Richardson J hews closely to the statute. He looks at the mechanics of the statute and concludes that the policy must be to allow groups that are groups on the last day of the year to consolidate. Richardson J' reasoning is recursive: [124]

> Certainly there appears to be no common pattern in international tax jurisprudence; and in New Zealand there have been substantial changes over the years reflecting differing legislative perspectives of the appropriate balance between the Revenue and the taxpaying community in this important respect. Thus the concepts of tax grouping and carry forward of losses employed in ss 191 and 188 respectively of the Income Tax Act 1976 must be characterised as tax concepts. They have no reality under the statute except in relation to income tax. In these circumstances it is difficult to discern any independent external yardstick of an overriding liability for income tax and the determination of the tax norms in this respect must turn on a close analysis of the specific provisions.

In autopoietic theory terms Richardson J is clear that determination of tax norms is a task for legal programmes, in particular by legislation Any "cognitive openness" to the environment is managed by the court in a way that maintains the closure of the legal system. The final sentence in the passage just quoted is a very clear example of recursive reasoning. For Richardson J the allocation of legal code values during structural coupling is never under any real threat from other systems. On appeal, however, the majority of the Privy Council viewed the question to be answered by the court quite differently, formulating it as:[125]

> Whether, as the Commissioner contends, s 191 takes effect as if the agreement did not exist, or whether, as Challenge contends, s 191 takes effect as if s 99 did not exist.

Framing the issue in these terms indicates the purposive approach that the Judicial Committee intended to take to the general anti-avoidance rule of section 99. This approach broadens the ambit of the enquiry beyond an analysis of statutory provisions that are in *prima facie* conflict. In autopoietic terms this approach positions general anti-avoidance rules to

[124] Ibid 552, line 49.
[125] Ibid 513, 556 per Lord Templeman

provide a structural coupling for greater "observation" of other systems with an interest in the tax event. In legal parlance it is a substance over form approach that general anti-avoidance rules necessarily mandate. As a result of his broader approach, Lord Templeman takes a different view from Richardson J of the policy behind the statute. He sees the policy of the consolidation rules as being to allow existing groups to consolidate. For the majority his Lordship held:[126]

> Section 191 was intended to give effect to the reality of group profits and losses. When one member of a group makes a profit of $5.8 million and another member of a group makes a loss of $5.8 million then the reality is that the group has made neither a profit nor a loss and that the members of the group should not be liable to tax. Section 191 in these circumstances is not an instrument of avoidance. But in the present circumstances the reality is that the Challenge group never made a loss of $5.8 million. A loss of $5.8 million was made by Perth and that loss fell on Merbank before Challenge contracted to buy Perth. Section 191 in these circumstances is an instrument of tax avoidance, which falls foul of s 99.

The phrases "give effect to the reality" and "the reality is that" reveal that the court intends to construct the "fact" of the transaction according to economic norms. This is the approach that is necessary to fulfil the purposive nature of general anti-avoidance rules and to answer the legal question as to whether the transaction was "avoidance". The economic reality, that is, the substance of the transaction as classified by the *economic* system informs the legal decision as to the code value that attaches to the transaction. The following passage emphasises the point:[127]

> In an arrangement of tax avoidance the financial position of the taxpayer is unaffected (save for the costs of devising and implementing the arrangement) and by the arrangement the taxpayer seeks to obtain a tax advantage without suffering that reduction in income, loss or expenditure which other taxpayers suffer and which Parliament intended to be suffered by any taxpayer qualifying for a reduction in his liability to tax.

---

[126] Ibid 558, line 33.
[127] Ibid 562, line 10.

Resolving legal questions by examining the scheme and purpose of a statute according to its terms may be appropriate for statutory interpretation in general, but it can scarcely be appropriate when general anti-avoidance rules are in play. Richardson J states that his objective is to characterise the relationship between sections 99 and 191 that "best reflects the intention of Parliament". [128] What Richardson J is really looking at is the policy behind the statute. But, logically, one cannot discover the policy of a statute just by reading the statute; by definition policy is something different from and prior to the statute. This is particularly true in a section 99 case. If courts take the view that taxpayers' transactions can satisfy the policy of a statute by locating themselves squarely within the terms of that Act, then a general anti-avoidance rule can have no effect. The rule must operate, if at all, *despite* the fact that taxpayers satisfy other terms of the statute. That is its *raison d'etre*.[129]

For these reasons, Lord Templeman's reasoning may be preferred to Richardson J's. His Lordship descries that the economic-cum-political policy of the statute is to permit companies that are members of groups to consolidate their accounts. The reason is that where companies are in a single group, then their shareholders, being natural persons, should bear tax (indirectly, through the companies) only to the extent that the group as a whole has made a profit. Otherwise, these natural persons indirectly suffer tax on profits that do not exist. But to allow a group to set losses that were suffered by a company when that company was owned by different natural persons off against the profits of the group is contrary to the economic policy of the loss consolidation rules, albeit that this policy cannot be discovered by an analysis of the language that lies between the four corners of the scheme and purpose of the Act. As a purposive provision, a general anti-avoidance rule directs the court to go beyond the language of the Act.

In the present context, the significance of these reflections is that the purposive nature of the general anti-avoidance rules presents a challenge to self-referential reasoning. Purposive legislation proves to be a problem for Luhmann's autopoietic theory. As Baxter

[128] Ibid 549, line 34

[129] Eg, *Challenge Corporation v Commissioner of Inland Revenue* [1986] 2 NZLR 513, 534 per Woodhouse P dissenting but affirmed, ibid 556 ff PC. Cf *Higgins v Smith* (1940) 308 US 473, 476 – 477, and *Elmiger v Commissioner of Inland Revenue* [1966] NZLR 683, 687 – 688 SC, both cited [1986] 2 NZLR 513, 532 per Woodhouse P.

puts it[130] purposive programmes "state general goals and give the legal decision-maker discretion as to how to realise them." Baxter notes that these programmes "do not appear to be cast in the form of conditional programmes"[131]as Luhmann views all legal programmes. Luhmann's theory maintains that programmes of the legal system are always conditional programmes whose role is to establish a combination of openness and closure.[132] Luhmann states:[133]

> Only conditional programmes can instruct the continuous linking of self-reference and external reference; only conditional programmes provide the system's orientation to and from its environment with a form which is cognitive and at the same time which can be evaluated deductively in the system.

As Baxter puts it, the effect is that:[134]

> The conditional program makes this combination [openness and closure] explicit, referring in its 'then' clause to the system-internal code and in its 'if' clause to the relevant event or process in the environment.

The present authors submit that Baxter is correct, but in being right Baxter reveals an important flaw in Luhmann's theory. If one were to try to show that a system is "operatively open" by reference to an example of deductive reasoningwith major and minor premises and a conclusion, a simple combination of "if" and "then" clauses would seem to be effective to make the point. That is, the example that Luhmann takes to demonstrate closure for the present authors would demonstrate openness. Luhmann argues[135] that there are no genuine purposive programmes and that programmes described in as purposive are better understood as variants of the conditional programme, where "the condition on which the choice of code

---

[130] Baxter, Hugh, "Autopoiesis and the 'Relative Autonomy' of Law" (1998) 19 Cardozo L. Rev. 1987, 2012.
[131] Ibid 2012.
[132] Idem.
[133] Luhmann, N, *Law As A Social System*, translated by Ziegert, K, (Oxford University Press, Oxford, 2004), 196.
[134] Baxter, Hugh, "Autopoiesis and the 'Relative Autonomy' of Law" (1998) 19 Cardozo L. Rev. 1987, 2012.
[135] Ibid 2013.

values depends is not a past fact, but a present projection of a future fact".[136] Luhmann, however, is not comfortable with this type of rule.[137]

> Purpose-specific programmes, however, do not allow for the setting of adequate limits to the facts that must be considered in a legal process.
>
> ...
>
> Orientation by purpose may well be a meaningful political perspective but as far as the legal system is concerned a lot speaks against it. On the one hand, the sensitivity of purpose-specific programmes to the conditions under which purposes can be achieved cannot be used to its fullest. On the other hand, purpose-specific programmes are, as far as legal technicalities are concerned, too imprecise to be able to block effectively any abuse or resistance to the achievement of their purposes.

For Luhmann autopoiesis is guaranteed not by a system's programmes but by its code. He therefore sees the real issue as being what the structural consequences are for the legal system, and for interpretations by other related systems in the social environment of the legal system, if incorporated purpose-specific programmes replaced the detailed conditioning of legal programmes.[138] To address this concern Luhmann posits[139] that "the conditional programming of law does not preclude that the purpose-specific programmes of other systems use law", however he acknowledges[140] that the systems and their types of programming must remain separate; only then can their interplay be productive.

This consideration prompts a further criticism of Luhmann's theory. The very distinction that Luhmann draws between coding and programming and the interplay between different systems' purpose-specific programmes may appear credible in the abstract but are readily challenged in practice. The distinction prompts inquiry into identifying what, if any, difference, exists, between coding and programming (perhaps better the subject of another paper) which in turn has consequences for the "cognitively open", "operatively closed" view of systems.

---

[136] Idem.

[137] Luhmann, N, *Law As A Social System*, translated by Ziegert, K, (Oxford University Press, Oxford, 2004), 196.

[138] Ibid 202.

[139] Idem.

[140] Ibid 203.

It is understandable, then, why Luhmann expresses concern for the adequacy or otherwise of judicial decision-making in purpose-specific programmes as posing a threat to the legal system. Of particular concern are future-oriented purposive programmes such as "best interests of the child". Baxter explains:[141]

> [Luhmann] takes a skeptical view of the legal system's capacity to forecast the future, suggests that considerations of purpose come dangerously close to overstepping the bounds of the judicial role, and worries that the demonstrated inaccuracy of judicial predictions undermines the authority of legal decisions.

Luhmann's concern about purpose-specific reasons in judicial decisions is that they expose judges to criticism, leaving only the authority of office and the necessity of decision-making to render judicial decisions valid.[142] Luhmann maintains that this does not, however, threaten the autopoiesis of the legal system "because it would be clear whom one should observe if one wanted to find out what is legal and what is illegal."[143] This does not seem an adequate basis on which to assert that the legal system still remains autopoietic.

It would appear that the substance over form approach authorised by general anti-avoidance rules in cases involving avoidance is more than a structural coupling for the purpose of "observation". Rather the coupling is between systems that are porous. An exchange occurs whereby the construction of the "fact" of the transaction is according to principles of the economic system. These principles determine the legal code value. To this extent the ectopia present in tax law that demonstrates the credibility of autopoiesis theory is weakened, and general anti-avoidance rules erode the self-referential, recursive nature of tax law by requiring it to accept a "fact" as constructed according to economic principles.

When courts deal with general anti-avoidance rules their role as paradox managers becomes critical. In the *Challenge* case, the contrasting approaches of Richardson J and Lord Templeman, illustrate how structural coupling by means of general anti-avoidance rules may

---

[141] Baxter, Hugh, "Autopoiesis and the 'Relative Autonomy' of Law" (1998) 19 Cardozo L. Rev. 1987, 2013.
[142] Luhmann, N, *Law As A Social System*, translated by Ziegert, K, (Oxford University Press, Oxford, 2004), 202.
[143] Ibid 202.

either maintain or breach the closed circle of tax law. The authors submit that whenever courts look to economic reality to inform the legal code value there has been an exchange between the legal system and the economic system such that the legal system is not just "cognitively open" but it is also "operatively open" to the economic system. In Lord Templeman's opinion, and no doubt in the opinion of Members of Parliament who turn their minds to the matter, that is the objective of general anti-avoidance rules. To adopt recursive reasoning in the face of a general anti-avoidance rule in the manner of Barker J and Richardson J in the *Challenge* case is to deny effect to the rule.

Summary and Conclusion

Autopoietic theory as presented by Niklas Luhmann is complex and highly abstract. Luhmann maintains that the basic units of a legal system are the communicative events that change legal structures. Legal communications, according to Luhmann, are circular and self-referencing. In this way the system is "operatively closed", establishing its boundaries internally, information being system-specific to the legal system that produced it. Nevertheless, Luhmann allows that through the mechanism of structural coupling legal system is able to "observe" other systems, which enables it to be "cognitively open" while at the same time remaining operatively closed.

Income tax law, being dislocated from its subject matter, presents an almost textbook example of autopoiesis. This dislocation or ectopia comes about because income tax law assesses a legalistic simulacrum of people's transactions, not the transactions themselves. A second reason for the ectopia of income tax law is that it is based on a number counter-factual fictions and counter-legal fictions.

In contrast, general anti-avoidance rules significantly erode the autopoietic nature of tax law. In Luhmann's terms, general anti-avoidance rules are programmes of the legal system that facilitate structural coupling with other systems, in particular the economic and political systems. However, the substance over form approach required by general anti-avoidance rules reveal a more porous boundary between systems than Luhmann would accept as possible; indeed, contrary to Luhmann, general anti-avoidance rules make income tax law

pro tanto not only cognitively open but also operatively so. Tax law is autopoietic in nature but ceases to be so when the general anti-avoidance rules are engaged.

Bird, Graham "Kant, Immanuel" in Honderich, Ted, *The Oxford Companion to Philosophy* (1995) 435, 436-437.

The Categorical Imperative

Kant's moral theory centres around the categorical imperative: "Act only on that maxim which you can at the same time will to be a universal law". Maxims are the general rules or principles on which rational agents act and they reflect the end that an agent has in view in choosing actions of a certain type in given circumstances. Thus, maxims are principles of the form: When in an *S*-type situation, act in an *A*-type manner in order to attain end-*E*. For example, I might make it my maxim always to pay my debts as soon as possible so as to avoid incurring unnecessary obligations. The categorical imperative tests maxims by prescribing a thought experiment in which one asks oneself whether one could consistently will one's maxim as a universal law, that is, one on which all other agents would also choose to act. The idea is to determine not simply whether the imagined universal law is consistent with itself, but whether its universal adoption is consistent with the agent's own ends and, therefore, something that the agent could consistently will. A maxim which passes this test is morally permissible, whereas one which does not is forbidden. Consider the maxim of borrowing money by falsely promising that one will repay. This maxim, Kant argues, conflicts with itself when universalized because it assumes a state of affairs in which promises to repay would not be believed and, therefore, the agent's project of profiting by false promising could not succeed. Consequently, policies such as false promising succeed only in so far as they are not universally adopted, so that in choosing them one makes an exception of oneself to a rule that one wills to hold for others.

The whole issue of the categorical imperative is extremely controversial, however, and there are a large number of interpretations and objections in the literature. The basic problem is that the test seems to yield both false positives such as "I shall smother infants who keep me awake at night by crying", which is clearly immoral but does not seem to be ruled out by the test, and false negatives such as "I shall play tennis on Sunday mornings when courts are available since everyone else is in Church", which seems both to fail the test and to be morally permissible. Although there have been many attempts to deal with these problems, it is not clear that any has been entirely satisfactory.

Blackwell Companions to Philosophy

This benchmark student reference series offers a comprehensive survey of philosophy as a whole. Written by many of today's leading figures, each volume provides lucid and engaging coverage of the key figures, terms, and movements of the main subdisciplines of philosophy. Each essay is fully cross-referenced and supported by a selected bibliography. Taken together, it provides the ideal basis for course use and an invaluable work of reference.

1. A Companion to Ethics
   Edited by Peter Singer
2. A Companion to Aesthetics
   Edited by David Cooper
3. A Companion to Epistemology
   Edited by Jonathan Dancy and Ernest Sosa
4. A Companion to Contemporary Political Philosophy
   Edited by Robert E. Goodin and Philip Pettit
5. A Companion to Philosophy of Mind
   Edited by Samuel Guttenplan
6. A Companion to Metaphysics
   Edited by Jaegwon Kim and Ernest Sosa
7. A Companion to Philosophy of Law and Legal Theory
   Edited by Dennis Patterson

Blackwell Companions to Philosophy

A Companion to Philosophy of Law and Legal Theory

Edited by
DENNIS PATTERSON

A Companion
and Legal The
Edited by Dennis

This outstanding vol
students and research
theory – a comprehen
the theories, topics, s
feature in the law scho
curricula.

Written by an interna
scholars, each of the 4
takes the reader throu
problems, theories, an
theory. While the text
uninitiated, the author
original and substantia

Taken as a whole, the v
work of reference for t
philosophy of law, juris
studies.

Copyright © Blackwell Publishers Ltd, 1996

First published 1996

2 4 6 8 10 9 7 5 3 1

Blackwell Publishers Inc
238 Main Street
Cambridge, Massachusetts 02142,
USA

Blackwell Publishers Ltd
108 Cowley Road
Oxford OX4 1JF
UK

All rights reserved. Except for the quotation of short passages for the purposes of criticism and review, no part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher.

Except in the United States of America, this book is sold subject to the condition that it shall not, by way of trade or otherwise, be lent, resold, hired out, or otherwise circulated without the publisher's prior consent in any form of binding or cover other than that in which it is published and without a similar condition including this condition being imposed on the subsequent purchaser.

Library of Congress Cataloging-in-Publication Data

A companion to philosophy of law and legal theory / [edited by]
Dennis Patterson.
p. cm. — (Blackwell companions to philosophy)
Includes bibliographical references and index.
ISBN 1-55786-535-3 (alk. paper)
1. Law—Philosophy. 2. Jurisprudence—United States.
I. Patterson, Dennis M. (Dennis Michael), 1955– . II. Series.
KF379.C66 1996
340'.1—dc20 95-46024
CIP

British Library Cataloguing in Publication Data

A CIP catalogue record for this book is available from the British Library.

Typeset in Photina on $10\frac{1}{2}/12\frac{1}{2}$ pt
by Graphicraft Typesetters Ltd, Hong Kong
Printed and bound in Great Britain by T. J. Press (Padstow) Ltd.

This book is printed on acid-free paper

Contents

List of contributors ix

Preface xii

PART I: AREAS OF LAW

1 **Property law** 3
JEREMY WALDRON

2 **Contract** 24
PETER BENSON

3 **Tort law** 57
STEPHEN R. PERRY

4 **Criminal law** 80
LEO KATZ

5 **Public international law** 96
PHILIP BOBBITT

6 **Constitutional law and religion** 113
PERRY DANE

7 **Constitutional law and interpretation** 126
PHILIP BOBBITT

8 **Constitutional law and privacy** 139
ANITA L. ALLEN

9 **Constitutional law and equality** 156
MAIMON SCHWARZSCHILD

Pettit, P. 1988: The paradox of loyalty. *American Philosophical Quarterly*, 25, 163.
Royce, J. [1908] 1920: *The Philosophy of Loyalty*. New York: Macmillan.
Sandel, M. 1982: *Liberalism and the Limits of Justice*. Cambridge and New York: Cambridge University Press, 15.
Schar, 1973: The case for patriotism. *American Review*, 17, 59.
Wolff, R. P. 1968: Loyalty. In *The Poverty of Liberalism*, ch. 2.

38

Coherence

KEN KRESS

An idea or theory is coherent if it hangs or fits together, if its parts are mutually supportive, if it is intelligible, if it flows from or expresses a single, unified viewpoint. An idea or theory is incoherent if it is unintelligible, inconsistent, ad hoc, fragmented, disjointed, or contains thoughts that are unrelated to and do not support one another. Historical, strict, rationalist, idealist coherence theories flow from a single principle. Modern *normative* coherence theories tend to be pluralistic.

After roughly characterizing seven properties which might enhance coherence, or be thought necessary or sufficient for it, this essay will set coherence theories of law in context by briefly describing coherence theories of truth, justified belief, ethics, and justice. Coherence of theories of *law* are then analyzed by asking: (1) what besides constitutions, statutes and precedents are in the base of legal sources; and (2) how that base is to be coherently reconstructed into valid law. Areas of agreement and disagreement among coherence theories are described. Representative coherence and (non-coherence) theories of law are examined in light of the above analysis.

Attention is next focussed on the concept of coherence itself. Three candidates for necessary requirements for coherence – consistency, comprehensiveness, the right answer thesis – are acknowledged to enhance coherence despite not being, at least generally, required for it. Seven techniques for enhancing coherence by eliminating or resolving conflicts among principles and counter-principles are described. The core concepts of coherence, monism, and unity, are then examined. A taxonomy of monisms and unities is developed which employs the seven techniques to characterize degrees of coherence in normative theories. Finally, the essay turns to the issue of the normative value of coherence and rejects Dworkin's claim that a coherent legal system is morally more legitimate than its less coherent counterparts.

As a first approximation, whether a theory is coherent, and if so, to what degree, may be analyzed under seven properties. Each property may be argued to be necessary for, or sufficient for, coherence. Alternatively, it may be claimed that the more of that property a theory manifests, the more coherent that theory is.

1 *Consistency*. A theory is consistent if its principles and propositions are logically consistent. In moral or legal theory, logical consistency is a weak constraint.

A stronger constraint would find a theory consistent only if its underlying principles are consistently applied in creating rules and deciding concrete cases.

2 *Comprehensiveness.* A theory is comprehensive if it provides answers (including the answer, "the theory cannot resolve this issue,") to all questions within the scope of the theory.
3 *Completeness.* A theory is complete if it provides single right answers to all questions within its scope, with no gaps, no unresolvable issues. Each proposition of a complete theory is either true or false, with no indeterminate or other truth values.
4 *Monism.* A theory is monistic if it flows from a single principle. It is nearly monistic if it flows from a handful of principles with a unified spirit.
5 *Unity* (*internal relations*). A theory displays unity when its principles imply, justify, or mutually support one another.
6 *Articulateness.* A theory is articulate if its methods for deciding issues, integrating and unifying its principles, and resolving conflicts among competing principles are expressed in language and are not merely "intuitive" techniques.
7 *Justified.* A theory is "justified" if it resolves conflicts with reasons. A pluralistic, normative theory is justified if its articulated meta-principles and means for resolving conflicts among principles are normatively intelligible.

Some would urge that utilitarianism is coherent because it is monistic and flows from a single normative principle: maximize happiness. Similarly, Sir Robert Filmer's Divine Right of Kings justifies a king's authority based on ancestry. Classical Newtonian mechanics manifests coherence because it is founded in three axioms of motion expressing a unified theory. Thermodynamics is similarly founded in a handful of principles expressing a single spirit.

Suggesting examples of seriously held incoherent theories is a dangerous occupation. Its proponents are likely to accuse one of bias and of failing to appreciate the simplifying and unifying aspects of the theory. A science in crisis, as described by Kuhn, would be incoherent insofar as it requires ad hoc principles to explain anomalies. The Ptolemaic, geocentric view of the universe is arguably incoherent because it employs ad hoc, unmotivated epicycles to describe planetary motion. The Copernican heliocentric theory is more coherent because it requires fewer, more motivated, and unified principles. Newtonian mechanics just prior to the discovery of relativity theory may be another example of a science in crisis, requiring ad hoc adjustments to explain anomalies. In normative theory, nihilistic and intuitionist perspectives are less coherent than those of their more articulated, justified, and optimistic opponents.

For example, Rawls criticizes intuitionists' failure to provide articulated, ethically justified metaprinciples which weigh principles off against counterprinciples because it ends rational discourse about normative matters prematurely: "An intuitionist conception of justice is, one might say, but half a conception" (Rawls, 1971, p. 41).

Coherence theories of law and morals may be understood as an outgrowth of coherence theories of truth and justified belief, which have a longer ancestry. The

coherence theory of truth can be crudely characterized as the view that a proposition is true if and only if it fits with other believed propositions. This theory sharply contrasts with the correspondence theory, which holds that a proposition is true if and only if it corresponds to the facts, where the facts are conceived as external to us, and independent of our beliefs about them (unless the proposition itself is about our beliefs). Correspondence theories of truth evoke the image of true propositions mirroring reality. The correspondence theory of truth is frequently associated with realism, the metaphysical claim that there exists an external world independent of our beliefs or social conventions. The denial of realism, anti-realism, which maintains that reality is dependent upon beliefs and conventions, entails the coherence theory of truth and may be entailed by it.

In modern theories of knowledge, coherence theories contrast with foundationalist theories. Foundationalist theories assert that certain basic beliefs – such as those reporting immediate sense impressions – are justified despite not being supported by (or inferred from) any other beliefs. Foundationalism maintains that justified beliefs consist of basic beliefs and justified inferences from basic beliefs. By contrast, coherence theories of justified beliefs maintain that a belief is justified if and only if it fits with other believed propositions. Foundationalist theories of justified belief conjure the picture of linear justification from basic beliefs; by contrast, coherence theories of truth and justified belief suggest a spider's web, a double geodesic dome, a link necklace, or a unified field. Perhaps the most illuminating metaphor for coherence is a puzzle with identically shaped pieces, say, one-inch squares, which must be arranged into a meaningful, coherent picture.

The most famous coherence methodology in modern normative theory is the technique of reflective equilibrium developed by Rawls to resolve issues about ethics and justice. Rawls's early method in *Outline of a Decision Procedure for Ethics* (1951) requires that we determine the considered judgments about concrete normative issues that would be made by individuals with average intelligence in idealized circumstances which promote integrity, impartiality, and insight. One then induces principles much like a scientist would – only the data to be explained are the morally competent actors' considered judgments, not scientific observations.

In *A Theory of Justice* (1971), Rawls substantially extends the method. Considered convictions no longer have epistemic priority over principles. We compare our considered concrete convictions – now about justice – to the abstract principles of justice chosen by rational individuals in circumstances designed to eliminate bias and self-interest to determine whether the principles and convictions fit together. There is an appropriate fit if someone following the principles would reach the convictions or, in the alternative, if the convictions could be viewed as a normatively attractive extension of the principles. If not, the method of reflective equilibrium engages in real work. Insofar as there are discrepancies between the principles and convictions, one or the other (or both) must be revised. The antifoundational aspect of the method consists in giving neither principles nor convictions *a priori* preference in the process of revision. We go back and forth revising first one and then the other until the principles imply the convictions, thereby rendering our convictions coherent and justified (Rawls, 1971, pp. 20–1).

What motivates coherence theories, especially normative coherence theories? Coherence seems desirable – or necessary – in a theory because what is coherent is intelligible and forms a rational, understandable unity rather than a patchwork quilt. Second, it appears that truth, morality, justice (and perhaps justified belief) must be coherent. In law, coherence accounts appear preferable to those legal positivist perspectives which link law to the intentions of their authoritative authors because it frees law from the dead hand of the past, promoting responsiveness to contemporaneous concerns (Raz, 1992, p. 292).

Each of these motivations for coherence is problematic. First, intelligibility does not guarantee truth, justification, or legal validity (but see Weinrib, 1988, 1995). Moreover, "coherence" as employed here is a technical philosophical term continuous with, but not identical to, its ordinary meaning (Raz, 1992, pp. 276–7). Second, as positivists and critical legal scholars urge, law may well be a patchwork quilt, the handiwork of political forces and actors proceeding at cross purposes with inconsistent ideologies. Third, while pure morality and justice must be coherent, given human fallibility, it is controversial whether practical day-to-day theories of morality, justice, and law should be coherent (Sayre-McCord, 1985, pp. 181–7). Moreover, while pure morality and justice must be coherent, surely positive law need not be. On the other hand, positive law requires techniques to eliminate strict inconsistencies. Why not limit incoherence by the same or similar methods?

Irrespective of the plausibility of other coherence theories, several considerations suggest that coherence theories of *law* have a special claim on us. The idea that law is a seamless web, that it is holistic, that precedents have a gravitational force throughout the law, that argument by analogy has an especial significance in law, and the principle that all are equal under the law, provide strong *prima facie* support for a coherence theory of law.

Coherence theories of law

Despite the apparent claim of some that a coherence theory of law is a coherence theory of truth for law (Kress, 1985, pp. 369–71; Raz, 1992, p. 283 and passim), a coherence theory of law is logically compatible with a correspondence theory of truth (Fumerton, 1994, pp. 90–1) or a deflationist theory of truth (Coleman, 1995, pp. 54–61). Theories of truth and theories of valid law are logically distinct, for all we know. For example, correspondence theories must be able to account for truths involving coherence, including claims that coherence is a (or the) determinant of legal validity. Consider the proposition that comparative negligence is valid law because it coheres better with general negligence principles than any alternative (including contributory negligence). That proposition will be true according to the correspondence theory just in case that proposition corresponds to the facts. The relevant factual questions are: first, does comparative negligence cohere better with negligence principles than contributory negligence or other alternatives and, second, assuming it does cohere better, does that mean it is valid law? If both

questions are answered affirmatively, the correspondence theory will declare the proposition "comparative negligence is valid law" true by virtue of *correspondence with "facts"* about the *coherence* of comparative negligence and general negligence, and about coherence as the criterion for legal validity. No persuasive argument has yet tied theories of law in any interesting way to general theories of truth.

Characterizing theories of law and adjudication clarifies differences among versions of coherence and non-coherence theories of law. With the exception of pure coherence natural law theories which assert that law is morality and justice, and that morality and justice are constituted by coherence, all coherence theories of law contain at least one non-coherentist aspect, and some contain more than one. This non-coherentist aspect consists in what Raz calls "a base" which is to be made coherent by some reconstructive method (Raz, 1992, p. 284). Almost all agree that bases include Constitutions, statutes, and precedents. Controversy arises over: (1) the best characterization of Constitutions, statutes and precedents; (2)(a) what else the base of a possible legal system could include, (2)(b) what else the base of some particular legal system does include; and (3) what method of coherent reconstruction (a) could in possible legal systems, or (b) does in some actual jurisdiction, produce valid law as output, that is, in what the coherence relation consists. Analyzing differences between coherence theories in these three ways is often useful, but the classification will not bear intense scrutiny. The categories are not entirely distinct. For example, natural law coherence theorists may be conceived as adding morality to the base, or as employing a method of reconstruction combining moral and coherence considerations.

Disagreements will arise, for example, about the best characterization of enactments and precedents. Does the base contain only present institutional acts or also past institutional acts? If so, how far back? Is there a principle of desuetude for statutes? Does the base include hypothetical acts which courts, but not legislatures, are prepared to make? Are hypothetical acts better accounted for under the method of reconstruction? Are precedents the words in reporters? The outcomes of cases given the actual facts? The outcomes given the facts as described by the court, plus the *ratio decidendi* and justifying principle provided by the court? Are dissenting opinions part of precedents?

Controversy exists over what besides enactments and precedents is in the base. Does it include conventional or critical morality? If the base includes conventional morality, is it constituted by judges', the legal profession's, or society's moral views? If the base includes conventional or critical morality, yet there is no foolproof method for determining what morality requires, what methods may courts legitimately employ to discover moral principles?

As developed in greater detail below, theories differ over whether reconstruction aims at coherence alone, or also at moral and political values. If other values are included, how are they combined with coherence? The best combination of coherence and morality? The morally best reasonably coherent theory? The most coherent, reasonably moral theory? *The most important differences are over what coherence itself means.* Is a theory coherent if its principles imply the base of legal sources? Justify the sources? If the principles and sources form a unified theory? If

the base is derivable from a single principle (monism)? If the theory provides an answer to all possible legal questions (comprehensiveness and right answer thesis?)

Non-legal coherence theories also require bases. In moral theory and the theory of justice, employment of an individual or society's moral beliefs as a base to be coherently reconstructed by reflective equilibrium, as in Rawls, appears ineffective. Unless it is reconstructed according to a justified coherence theory of truth, the reconstructed set appears as likely to reflect biases and socialization as moral truth.

Employment of a base of beliefs in coherence theories of truth and justified belief may illegitimately sneak in a realist or foundationalist element. By contrast, employment of a base of legal sources is innocuous. It serves to secure a coherence theory, or any legal theory, to authoritative sources of law (Raz, 1992, p. 291 and n. 29). Only extreme natural law perspectives might question the legitimacy of a base in law.

Raz's parsimonious positivist sources thesis (1979, ch. 3, 1985) maintains that law consists of source-based law only: constitutions, statutes, and precedents. Since legal authorities may generate an incoherent mishmash of legal sources, the sources thesis is not a coherence theory.

A positivist pure coherence theory maintains that, in addition to the base, law includes those principles and policies which cohere with – by implying – the base. Finally, anything which follows from the principles, policies, and the base is law (cf. Sartorius, 1975, p. 192). This theory parallels Rawls's early theory of reflective equilibrium (1951, described above) with the base of enactments and precedents playing the role which considered convictions play for Rawls. Like considered convictions in Rawls's early work, enactments, precedents and other sources of law have priority over abstract principles (and policies) – the principles are chosen to fit (by implying) the sources of law. The coherence in this theory is a version of unity. The underlying principles and policies imply the base and are in that way internally related to it.

Quine's under-determination thesis implies that many sets of principles and policies will imply the base. This suggests that where coherence is understood as the principles implying the base many sets of principles will be equally coherent because they each imply the base. Yet the theory supplies no method for choosing among those multiple sets of principles and policies. To minimize indeterminacy, coherence theories might be motivated to employ non-coherentist elements such as morality as a tie breaker between equally coherent theories. That is, morality could be a second element in a lexical ordering (defined below).

Other coherence theories may define law as the best combination of coherentist and other considerations. Coherence is one element among many. For example, law consists of the best combination of unity, comprehensiveness (coherentist elements) and morality (a noncoherentist element). Where coherence is one factor among many, however, theories of law may be incommensurate and indeterminate. One theory is more unified and comprehensive, another is morally preferable. Unless there is a metric which balances these values against one another, neither theory is better than the other. Nor are they equally good. In mathematical

jargon, this coherence theory is only partially ordered. Dworkin chooses a yet more complex relation between coherentist and non-coherentist aspects: law is the morally most appealing of all those sets of principles and policies which explain or imply the legal base.

Burton's positivist theory (1985; 1995) takes a more expansive view of the sources of law, including within it dissenting opinions, legal scholarship, and conventional moral beliefs within the legal profession, in addition to constitutions, statutes and precedents. Law, for Burton, is the most coherent reconstruction of the legal community's beliefs and dispositions about what would lead to order and justice.

Eisenberg's positivist theory of the common law (1988) is worth exploring at greater length than can be provided here. His discussions of overruling (pp. 104–45) and of what constitutes legally acceptable evidence that a proposition has social support (pp. 16–19, 29–32, 40–2) are especially valuable contributions, and deserve more attention than they have received. Eisenberg begins with a base of sources similar to Burton's, although he looks beyond the legal profession to society at large to determine conventional moral norms and policies.

Eisenberg argues that conflicts among moral norms may be resolved by reflective equilibrium, on the basis of which norm fits better with policies, or determining which norm fits better with doctrinal propositions. The theory provides for two more important roles for coherence as a regulative ideal. First, the social propositions should imply the valid doctrinal rules. Second, the body of valid legal rules should be consistent in the sense that the principles, policies, and the like which imply those doctrinal rules must be consistently applied. Although in ideal theory doctrinal propositions would reflect social propositions perfectly, and be as close to being implied by them as is possible in normative practices, in the real world, this will not be so. Eisenberg adds an overarching non-coherentist element: doctrinal propositions will and should lag behind changes in social propositions in consequence of rule of law and other conservative principles which ground a principle of doctrinal stability, thereby slowing the evolution of doctrinal propositions toward conventional morality and policy.

The most famous and influential coherence theory of law and adjudication, Dworkin's natural law theory, is discussed below. A pure natural law theory holding that law is critical morality, political theory, or justice is coherentist to the degree – if any – that its theory of morality, political theory, or justice is coherentist, and not otherwise.

What coherence is

Beyond providing a base, coherence theories must explain how to modify the base to produce law as output. Put differently, a coherence theory must specify in more detail what it means for legal norms to cohere or fit together. Coherence theorists agree that some – but not extensive – modification of the base is permitted.

To understand coherence in law, techniques thought to promote coherence, or

properties or states thought to be aspects of, explanations of, or to be necessary or sufficient for coherence, will be examined. The discussion will focus on justificatory coherence within normative theory, particularly ethics and law, although concepts more appropriate to the theory of knowledge will be discussed in passing. The primary aim throughout is to serve as background for the later taxonomy of coherence in normative theory, although much of what is said here applies more generally.

First, three possible necessary requirements for coherence – consistency, comprehensiveness, and the right answer thesis – will be described and claimed to generally enhance coherence. Still, comprehensiveness and the right answer thesis will be found not to be necessary for coherence, while consistency is necessary in ideal theory but not in practice.

Second, seven techniques aimed at reducing or resolving conflicts among principles of theories, such as pre-emption and reflective equilibrium (described above) will be examined. Characterized thinly as bare or mechanical methods for resolving conflicts, the techniques appear relevant only to consistency, comprehensiveness and the right-answer thesis dimensions of coherence. The appearance is misleading because the conflict-resolution techniques can be articulated and justified. For example, in the United States, federal law expressly or implicitly intended to cover a field pre-empts state law because: (1) the ultimate authority, the US Constitution, so provides; (2a) Congressional power to create uniform national law is essential or helpful to sound policy and moral ends, and (2b) state authority to develop law in the absence of Congressional pre-emption is also desirable. So articulated and justified, federal pre-emption of state law may be part of a coherent unity of doctrine exemplifying a unified spirit in which abstract principles of federalism imply (and are thereby internally related to) principles of pre-emption which in turn imply pre-emption rules and outcomes in concrete cases.

Finally, the core concepts of coherence – monism and unity (internal relations) – will be examined.

Consistency at a time is necessary *in theory* for normative coherence, but is not sufficient for it. By normative coherence is meant a coherence theory in a normative area where the coherence requirements do substantial justificatory work. *Consistency over time* is not necessary for a coherence theory such as Dworkin's, which permits – indeed requires – change over time. Insofar as common-law adjudication is one of the features to be explained by coherence, the theory should not demand consistency or coherence among the principles of the theory at different times, but only a coherent path of movement over time (Kress, 1985). Finally, although consistency at a time is necessary in theory, or metaphysically, for coherence, it is not necessarily required in practice. Although we aim for consistency in the long run, modest skepticism may recommend that we do better day-to-day if we retain some inconsistencies until we are able to resolve them satisfactorily, rather than force consistency via ad hoc solutions. Given the difficulty of developing consistent, coherent, and complete theories, and the value of experimentation, especially in a federal system, consistency (and coherence) are arguably less desirable and necessary *in practice* than as a regulative ideal (Sayre-McCord, 1985, pp.

181–7), despite Rawls and Dworkin's insistence that coherent explanations articulating underlying principles are required of governmental actors to minimize the prospects for bias, self-interest, and deceit, and thereby to help legitimize the use of force and coercion (for example, Dworkin 1978, pp. 162–3; but contrast, Dworkin, 1986, pp. 217–19).

Some claim that coherence requires that a theory must be *comprehensive* and cover the entirety of the relevant field, supplying an answer to each question within its scope, including, where appropriate, the answer "indeterminate." For example, a theory might hold abortion legal or moral in the first trimester, and illegal and immoral in the final trimester; but indeterminate in the middle trimester because at that stage of gestation it is indeterminate whether the fetus has a right to life. This perspective is unduly restrictive. Comprehensiveness improves coherence, but is not required for all conceptions of it.

More controversial yet is a third possible requirement of coherence, the *completeness* requirement. Called by some the "right answer thesis" and by others the "bivalence thesis," it maintains that each proposition within the scope of the theory is either true or false, with no gaps, no unanswerable questions, and no indeterminate truth values. A normative theory may be substantially coherent, even while leaving some vague, borderline or other cases unanswered. The right answer thesis is not necessary for coherence. Nonetheless, it cannot be denied that in certain circumstances, right answers will enhance coherence, while gaps will undermine it.

Shifting focus to methods for eliminating or resolving conflicts and deciding concrete cases, one coherentist method is *reflective equilibrium*, discussed above. A second coherentist technique weighs and balances norms against one another. *General equilibrium* – a third route to coherence – succeeds when things fit together even when individual elements are warring; an overall theory may make sense although its principles conflict (Dworkin, 1986, p. 183; Hobbes, 1962, pp. 105, 110, 164, 229; Sartorius, 1975).

A fourth coherentist method is *lexical ordering*. In a lexical ordering, the first principle must be completely satisfied before the second principle is considered; the second must be completely satisfied before the third is considered; and so on. In this way, Rawls asserts, a lexical ordering avoids weighing and balancing and gives earlier principles "absolute weight, so to speak, with respect to later ones." (1971, pp. 42–3, 60ff). Rawls ranks the principle of equal liberty before the difference principle distributing social and economic resources (1971, p. 61).

A fifth coherentist method is *scope*. By restricting some principles to mutually exclusive areas, concrete conflicts cannot arise. Each standard is limited to its own sphere of influence. For example, a jurisdiction might provide that the principle of equal opportunity governs official employment, while private employment is regulated by freedom of contract. Similarly, antidiscrimination principles might regulate government and public enterprises while freedom of association regulates private activities like private clubs.

Pre-emption, a sixth coherentist method, assists in avoiding conflict among standards. If occasionally, or always, when two principles conflict, one pre-empts the

other over all or some portion of their range, potential inconsistency and incoherence will be avoided. In the United States, federal statutes intended to cover a field pre-empt state law on the same subject, thus (in principle) avoiding conflicts.

In their arguments that law is indeterminate, incoherent, and contradictory, critical legal scholars cite the lack of explicit meta-principles to adjudicate among competing principles and counter-principles (Kennedy, 1976, pp. 1,723–4). A seventh possible way for a theory to achieve coherence is to encompass explicit *meta-principles* which resolve conflicts among principles.

The two most important aspects of coherence have been saved for last: *monism* and *internal relations*. The "single fountain" or monistic theory aims to avoid or resolve all conflicts by confining the theory to one fundamental principle from which all subprinciples follow. Utilitarianism is a well-known example of a single fountain theory.

In fact, monism does not guarantee the absence of concrete conflict. First, conjoining multiple principles into a "complex fundamental principle" will permit the conjoined principles to conflict as in any pluralistic theory, and there is no easy way to discriminate between true monistic theories and pluralistic theories in monistic dress. Even true monistic theories may yield conflicting directives under certain factual circumstances: "obey your parents" is the fundamental principle, but mother and father give inconsistent commands. Moreover, even when a monistic theory avoids conflicts, it might fail to provide right answers as a result of vagueness or a failure to be applicable under the circumstances.

Along with monism, the most important consideration in assessing coherence is the internal architecture within the theory – that is, the internal relations among the principles (and norms, rules, policies) of the theory. In its strictest version, internal relations (unity) require that each principle entail and be entailed by every other principle. Such strict versions of internal relations are implausible for most normative theory. Less strict versions of internal relations require that each principle entail or be entailed by some other principle or principles of the theory. Even less strict versions hold that each principle must justify or be justified by some other principle, explain or be explained by another principle, make probable or be made probable by another principle, or be evidence for or be supported by some other principle. These even less strict versions of internal relations may be called one-one versions, to distinguish them from the one-many or one-all versions. For example, one-all versions of internal relations require that each principle be related by some inferential, justificatory, or evidentiary relation not to a single other principle, but to all the rest as a whole. That is, the other principles, taken as a whole, entail, justify or support the principle in question.

Strong internal relations promote coherence by limiting the risk of multiple subject matter incoherence. An example of multiple subject matter incoherence is the following three sentences. Sally is smart. Two is the smallest prime number. Inflation is not a function of the money supply. These three sentences lack coherence because they are about too many different subject matters, and are not mutually supportive.

Another version of internal relations is narrative: I came. I saw. I conquered (Balkin, 1993, p. 114).

Strong versions of internal relations have been out of style for half a century. Weaker forms have survived – for instance, in Dworkin's requirement that the individual coherent principles underlying legal doctrine fit coherently with each other.

Internal relations as described above in either one–one or one–many versions do not guarantee coherence. Thus, it is possible that several subsystems of principles exist, where each member of each subsystem is related by the relevant inferential or justificatory relation to other members of that *same* subsystem, yet the different subsystems have at most weak or no relation to one another. In this way, internal architecture that relates principles one to one, or unidirectionally many to one, does not necessarily prevent multiple subject matter incoherence. Prevention of multiple subject matter incoherence is more likely if the justificatory–inferential relations among principles are reciprocal, holistic, and pervasive. The degree to which this obtains is one measure of the degree of coherence within a theory (Bonjour, 1985, pp. 97–8).

The characterization of normative coherence theories

From here on, the discussion is limited to coherence in moral, legal, and political theory, and would not necessarily apply to coherence theories of knowledge or truth. The remarks that follow are not intended to be a complete and final definition of coherence in normative theory. Coherence is much too difficult a concept for that. There is a range of conceptions of coherence, not just one. What is offered here is a first approximation of a taxonomy of conceptions of coherence.

Referring to reflective equilibrium and concluding the discussion is inadequate to account for coherence in normative theories. The analytical tools described thus far provide the means for a richer, deeper, and more comprehensive analysis.

Moreover, characterizing varying conceptions, and how they differ, will advance efforts to assess their normative virtues. Distinguishing the subconcepts of coherence permits identifying which are part of a particular conception of coherence, singly or in combination. Such identifications aid evaluation of whether a particular conception is morally legitimate, desirable, or partakes in any other normative virtue by determining whether the subconcepts it has induce those virtues.

As noted above, comprehensiveness and right answers enhance coherence but are not necessary for all forms of it. Consistency at a time is a theoretical but not a practical requirement for coherence. The core of coherence is monism and internal relations.

Three versions of monism shall be discussed. No claim is intended that other categorizations are unworthy of investigation. First, strict monism requires that the entire theory "flow from" one master principle so that each subprinciple is implicit in the master principle. "Flow from" means entailment, near entailment,

or similar logical and theoretical relations. Moreover, the subprinciples must flow from the master principle together and in harmony so that they are an integrated whole. Weinrib's formalist view that private law, especially tort law, is justified and made intelligible by the principle of corrective justice is a strict monism. Corrective justice holds that those wrongfully causing harm must compensate their victims. Weinrib (1989, 1995; see Article 21, LEGAL FORMALISM) claims that the principle of corrective justice explains, integrates, and harmonizes tort law's bipolar procedure, breach, causation, and damages requirements (Kress, 1993, pp. 648–9, 659–61).

Second, there is moderate monism, wherein methods of reflective equilibrium, weighing and balancing, general equilibrium, lexical ordering, scope, pre-emption, meta-principles, and the like resolve competition and conflict among principles and counter-principles, thus achieving substantial or complete coherence and consistency. Moreover, some master principle or norm explains for each of the following subprinciples that are employed in the theory:

1 why reflective equilibrium eliminates some data and principles, keeps the rest, and adds others;
2 why principles and counter-principles are balanced as they are, why each has the weight (in context) it does (and what justifies the particular weighing mechanism employed);
3 why general equilibrium reaches stasis and makes the theory intelligible;
4 why the lexical ordering is ordered as it is;
5 why the various norms are limited in scope as they are;
6 why those norms which pre-empt others do so;
7 why the particular meta-principles employed for resolving conflict are appropriate; and so on.

In short, the master principle provides a *normatively intelligible explanation and articulation* of arithmetical and abstract methods and meta-principles for resolving conflict among principles. Such a master principle, in combination with the resolution device, serves as the monistic principle, and as a functional substitute for *strict* monism. Yet it is not strict monism, because it allows for disharmony among subprinciples and for subprinciples that do not flow from the master principle. In near monism, a small number of principles irreducible to each other perform the function of the master principle.

By subtracting some articulateness and justifiedness, the second version of monism is transmuted into a third where resolution of concrete cases is accomplished via reflective equilibrium, weighing and balancing, general equilibrium, lexical ordering, scope, pre-emption, meta-principles, and the like, but without recourse to any articulated master principle. Nevertheless, the principles, norms, and conflict resolution devices must reflect a single, unified normative vision. This conception of monism is supported by pragmatist impulses and by atheoretical interpretations of Wittgenstein. Whether these are forms of monism is debatable.

If, however, the resolution devices are interpreted as creating functional monism (plausible, perhaps, if they resolve almost all issues) and the normative vision is clear, version three may be monistic.

Ironically, the first two versions of monism are foundationalist in the respect of being built upon a master principle (except for near monism). Modern coherence theorists might reject this foundationalist imputation, claiming that the master principle is induced by the scientific method of hypothetico-deduction, that is, by determining which master principle (best?) implies the lower norms, as in Rawls's early version of reflective equilibrium, described above. A better perspective simply rejects the alleged opposition between coherence and foundationalism in *law*. Although foundationalism contrasts with coherence theories of justified belief (although even here sophisticated versions of either incorporate aspects of the other), the contrast is not universal: coherence theories of truth are generally compared to realist, correspondence theories, not foundationalist perspectives.

The concept of unity is close to monism in spirit, but it is distinguishable in form, focussing more on internal architecture among the principles of the theory. One way to look at unity is through the version of internal architecture employed. That perspective is more appropriate to theories of justified belief and will not be discussed here. Instead, a taxonomy halfway between that and the classification provided for monism will be explored.

First, unity might be provided by a single master principle that entails all the principles, and thereby relates each to the others in a normatively intelligible fashion. In weaker forms of this first version of unity, the single master principle justifies, explains, makes probable, is evidence for, or otherwise supports all the rest of the principles.

Second, the principles might be united by some form of reflective equilibrium, weighing and balancing, general equilibrium, lexical ordering, scope, pre-emption, meta-principles, or the like that is entailed (or, in weaker versions, justified, explained, made probable, or otherwise supported) by an articulated master principle.

Third, unity might be achieved by one or more of the techniques of reflective equilibrium, weighing and balancing, general equilibrium, lexical ordering, scope, pre-emption, meta-principles, or the like, without any articulated supporting master principle. In its strongest form, this unity would be a matter of entailment between the principles, each to every other. In a weaker version, each would entail or be entailed by some other. Yet weaker versions, would replace entailment with justification, or probabilistic and similar less-strict evidentiary relations. Finally, some minimal unity may be achievable by the above techniques without any evidentiary relations being created (or existing?) among the principles of the theory. The third version of unity may be understood as the second version with less articulateness.

Strong versions of monism entail some version of unity or internal relations. Theories exemplifying weak forms of version three for both monism and unity are, at best, weakly coherent.

Consistency (in theory but not practice) and at least one of monism or unity are clearly necessary for ideal coherence. It is less certain whether both monism and

unity are necessary for coherence. But it is clear that, with consistency, they are sufficient for it. There is no single central concept of coherence, but instead many different conceptions of it. Stronger forms of monism and unity give rise to stronger versions of coherence.

The normative value of coherence

An important question about coherence which will only be discussed briefly here is whether a legal system exemplifying coherence, or a coherent legal (or normative) theory is more morally legitimate, desirable, or better respects individual rights than one which does not. Naturally, the answer will differ for different conceptions of coherence. Dworkin's claim that legal systems manifesting the version of coherence he dubs "integrity" better legitimizes law than less coherent legal systems will be evaluated, although the arguments employed are intended to apply broadly to other coherence theories.

Dworkin's early writings maintained that a legal proposition is true if it follows from that (coherent) scheme of principles which best justifies and explains the precedents, statutes and constitution. His conception of coherence was a version of Rawls's mature method of reflective equilibrium in *A Theory of Justice*, emphasizing the requirement that the underlying principles must be consistently applied in justifying surface rules and reaching concrete judicial decisions.

Dworkin's mature theory differs slightly in a way which, as noted earlier, admits a second non-coherentist element besides the base and obscures its connection to a pure coherence theory (Kress, 1985, p. 378, n. 53). In Dworkin's later writings, a proposition is law if it follows from the morally most appealing set of principles that meet or exceed a (vague) threshold of fit with legal institutional facts (constitutions, statutes, precedents (Dworkin, 1978, pp. 340–1, 360)). In yet later writings, he allows that the threshold is not an absolute floor: one may drop beneath the threshold for urgent or exceptional moral gains (Dworkin, 1986). Raz argues – somewhat disingenuously – that Dworkin may not be a coherence theorist because in *Law's Empire* the moral elements – justice, fairness, and due process – do all the work, leaving coherence (fit) idle (Raz, 1992, pp. 315–21, esp. 317).

Dworkin supported coherence in his early work with four main abstract arguments. The first three were arguments for his rights thesis that judges do and should decide cases on the basis of principle, not policy. The rights thesis imposes a constraint of coherence on judges since Dworkin defined principles as requiring an equality – consistency in application – much stricter than that required of actions justified on policy grounds.

The rights thesis was defended by Dworkin on three main abstract grounds. First, Dworkin argues that the democratic principle that elected and politically accountable officials, and not judges, should make law has far more force against judicial decisions generated by policy than it has against those generated by principle. The second abstract political argument for the rights thesis is that it is unfair for the judge to create a new duty based upon policy and apply it retroactively

because "then the losing party would be punished, not because he violated some duty he had, but rather a new duty created after the event" (Dworkin, 1978, p. 84; but see Kress, 1984). Third, Dworkin argues that the best explanation of the requirement of precedent that like cases be treated alike is that adjudication is restricted to arguments of principle because principle requires more consistency from case to case than policy does. In addition to the support provided for coherence by the argument for the rights thesis, Dworkin argues directly that the doctrine of political responsibility requires consistent, articulated rationales for government actions.

One objection to coherence theories of law, including Dworkin's, is that they are path-dependent (Kress, 1989, pp. 5–53, esp. 50–3, 20–6) and lead to morally troubling retroactive application of law. Dworkin's theory of legal reasoning is built on the proposition that litigants are entitled to the enforcement of pre-existing legal rights. One way in which Dworkin's early work expressed this major claim was by presupposing that litigants have a right to have decisions determined by settled law. In Dworkin's theory, settled law together with moral theory determines litigants' rights and litigants' rights determine the proper outcome. Thus, decisions are a function of, among other things, settled law.

Retroactivity is a consequence of legal rights being a function of settled law and upon the temporal gap between events being litigated and their eventual adjudication. Judicial decisions change the settled law. Often, if not always, the settled law will be changed between the occurrence of events being litigated and their eventual adjudication. In consequence, a litigant's rights will sometimes also be changed. If changes in the settled law change the dispositive legal right, the litigant who would have prevailed given the legal rights existing at the time of the occurrence will lose because she no longer has the right at the time of the adjudication. The opposite is true of the opposing litigant. This is retroactive application of law (Kress, 1985, 1989; Alexander and Kress, 1995, pp. 296–301).

Hurley (1990), following a suggestion of Dworkin's, objects that legal rights do not change in coherence theories when judges decide cases correctly, but only when judges make mistakes. This is a problematic conception of precedent. What could justify a doctrine of precedent which provides a reason to follow (and a right to equal treatment of) mistaken and unjust decisions, but no reason to follow correct, just ones? (Alexander and Kress, 1995, p. 300).

Raz (1985, 1992) claims that coherence theories are unable to satisfactorily account for the authority of law, and the proper role of legal sources within the law. Although Raz views the criticism as a structural, evaluative but not moral criticism, it could also be read as a moral critique (cf. Perry, 1995). Raz claims that, in general, legal authorities are legitimate to the extent that they aid individuals in doing the right thing, that is, in acting in accord with the reasons that apply to them (normal justification thesis). This requires that legal authorities promulgate norms on the basis of the reasons and circumstances which apply to individuals (dependence thesis). Where legal authorities are better than individuals at figuring out what the applicable reasons require of individuals, individuals following the authoritative acts of legal authorities will act correctly more frequently

than if they follow their own lights. But this means that once an alleged legal authority has been shown to be legitimate (generally by the normal justification thesis), then its authority is respected only if individuals to whom its decisions apply act on the basis of the authority's reasons and standards, and not on the basis of the reasons which applied to the individuals prior to the authority's utterance: authoritative utterances pre-empt dependent reasons (pre-emption thesis).

But coherence theories, including Dworkin's, are inconsistent with this conception of the authority of law. First, the most coherent account of the legal sources may be entirely original, thereby severing all connections with the reasons and norms uttered by legal authorities. Dworkin's conception cannot account for the mediating role of authority.

Moreover, it denies the pre-emption thesis. On Dworkin's theory, citizens deciding how to act must determine the morally best reconstruction of the authoritative sources of law. But this means that the identification of law depends upon the very considerations which applied to the individual before the authoritative act and which law is supposed to settle. In summary, "Coherence accounts take the base [of legal sources] because it is too absurd to disregard it; then they strive to ignore it and to explain the law in a way which transcends the inherent limitations of the workings of human institutions, and by transcending them they misunderstand them" (Raz, 1992, p. 297).

An urgent question about the moral value of coherence is whether it promotes the moral legitimacy of legal systems exemplifying it. By coherence, which he now calls "integrity," Dworkin means at least:

1(a) the principles underlying official government acts must be individually coherent and intelligible;
1(b) the individual principles must be consistently applied, with applicable principles receiving similar weight in relevantly similar situations;
2(a) the principles, as a whole, must be consistently applied, with like situations being treated alike;
2(b) the principles as a whole must fit together into a single and comprehensive vision of justice.

Moreover, the justification for each of (1)–(2) is the same: "consistency in principle . . . requires that the various standards governing the state's use of coercion against its citizens be consistent in the sense that they express a single and comprehensive vision of justice" (Dworkin, 1986, pp. 134, 88, 116–17, 166; Alexander and Kress, 1995, p. 311). The justification for coherence generally and of the specific conception of it embodied in (1)–(2) is importantly connected to the nearly universally accepted internal point of view that law is a matter of practical reason, intended to provide guides to conduct for its subjects, and grounds for criticism of violations of its commands. While an external observer might be able to understand a foreign legal system as a patchwork quilt of conflicting norms resulting from the respective fortunes of opposed ideologies, someone adopting an extreme

internal point of view cannot do so. She regards legal norms as valid and as guides to her behavior and judgment. Yet she cannot accept the norms as her own unless she can regard the norms as "valid and justified, and [she] cannot regard them as justified unless they form a coherent body" (Raz, 1992, p. 293; Dworkin, 1986, p. 189).

Dworkin gives several arguments explaining how the form of coherence exemplified in the theory he calls "law as integrity" legitimizes law as integrity. The argument that he develops at greatest length, and on which he relies most heavily, is the argument from community. To understand this argument, we must recall that for Dworkin, integrity involves discerning principles that underlie and justify governmental acts (such as legislation and judicial decisions) and following those principles in making future governmental decisions (such as decisions in hard cases at law). Dworkin claims that "[A] political society that accepts integrity as a political virtue becomes a special form of community, special in a way that promotes its moral authority to assume and deploy a monopoly of coercive force" (1986, p. 188).

Dworkin's argument for this proposition is difficult to decipher; what follows is a reconstruction. Dworkin begins with the claim that the major traditional grounds for political legitimacy and political obligation (consent, tacit consent, fair play, duty to uphold just institutions) do not succeed in bestowing legitimate authority on governments to coerce and use force; nor do they generate moral obligations on citizens to obey the law. An interpretive reconstruction of the argument from fair play, however, is sufficient. Dworkin's new interpretive version of the argument from fair play reconstrues political obligation as a form of associative obligation – that is, the obligations arising within groups or communities such as families, law faculties, and clubs. For Dworkin, a bare political community, such as the United States, is a true community giving rise to true associative obligations only if the obligations arising from the community have four characteristics, of which two are relevant here:

1 each must be thought of as flowing from an underlying and pervasive concern for the other members; and
2 each must be predicated not only on concern for the other members, but on equal and reciprocal concern.

Of particular concern here is the first requirement that particular obligations must be thought of as flowing from a deep, underlying, and pervasive concern for other members of the community, because Dworkin argues that a community that accepts integrity as political ideal is better able to meet this requirement than a community that does not. This first requirement tracks Dworkin's early criticism of legal positivism that besides rules the law includes those principles which underlie and justify the rules (1978, pp. 14–45).

Dworkin claims that citizens' political obligations clearly include those obligations laid down in explicit rules of law. To satisfy the first requirement, however, citizens' obligations cannot be thought of as exhausted by the explicit rules. Citizens

must think of themselves as having whatever obligations and rights can be shown to flow from the values of equal concern that underlie the explicit rules.

Citizens can infer such inexplicit obligations and rights from the values underlying the explicit only if the explicit rules are coherent. If the explicit rules are incoherent, the only principles and values that can be said to be underlying them will be similarly incoherent, or even contradictory. This incoherent or contradictory foundation would frustrate citizens' attempts to successfully infer, and engage in dialogue about, what their inexplicit obligations and rights are. If individual principles are incoherent or unintelligible, citizens will not be able to apply them. If the individual principles are not consistently applied, but receive different weights in similar situations, their weight in novel but similar situations will be indeterminate. If the principles as a whole are not consistently applied but principles and counter-principles with the same pattern of relative weights sometimes are decided in favor of the principles and sometimes in favor of the counter-principles, new situations with the identical pattern of opposing norms will be unpredictable and indeterminate. If the principles do not fit together into a single and comprehensive vision of justice, then they contain all, or part, of contradictory visions which cannot be reconciled, or they are incomplete or incoherent. Once again dialogue and decision making will be hindered or prevented altogether. Thus, the political system can be legitimate only if its explicit rules and the underlying principles are coherent.

There are many ready avenues of attack of this reconstruction of Dworkin's text. For example, the duty to uphold just institutions, in addition to the argument from fair play, may legitimize government. But let us ignore these issues.

For Dworkin, an act's manifesting integrity does not insure its justice and legitimize it. The relationship is more complicated than that. Rather, a governmental act justified by justice, charity, efficiency, or other grounds, is permitted and legitimate only if that act can be shown to flow from, or cohere with, other actions taken in the name of the community in the past (Dworkin, 1986, p. 93; Alexander and Kress, 1995, pp. 312–14). But this constraint of coherence is unattractive if it prevents a government which has so far limited itself to promoting the welfare of its citizens from broadening its horizons to doing justice, or offering its services as an international mediator, or engaging in charity, unless it can engage in intellectual contortions demonstrating that these new roles cohere with prior governmental acts.

More importantly, it is not true that when a new initiative or incoherent act is taken, this disrupts the ability of citizens to engage in dialogue about, or judges to make inferences about, what the law now is, thereby undermining law's legitimacy. For example, suppose that until now a jurisdiction has operated on a color blind antidiscrimination principle. Proponents of affirmative action now convince the legislature to vote, or a court to hold, that affirmative action for some historically discriminated against group is required. The defender of integrity will allege that the law has now become incoherent in a way that diminishes the ability of citizens and officials to coherently work out the law's implications. But the defender of affirmative action may respond that the new affirmative action law brings

along with its own set of underlying principles, on the basis of which dialogue and inference about the law's commands should proceed. There is no loss in the ability to engage in dialogue; rather, the principles justifying affirmative action have now replaced the principles of color blindness in making authoritative determinations. The Dworkinian conception of coherence does not legitimate law better than following the demands of morality and justice. (For a fuller development of the argument, see Alexander and Kress, 1995, pp. 308–26.)

Bibliography

Alexander, L. and Kress, K. 1995: Against legal principles. In A. Marmor (ed.), *Law and Interpretation*, Oxford: Clarendon Press, 279–327.
Balkin, J. 1993: Understanding legal understanding: the legal subject and the problem of legal coherence. *Yale Law Journal*, 103, 105–76.
Bonjour, L. 1985: *The Structure of Empirical Knowledge*. Cambridge: Harvard University Press.
Burton, S. 1985: *An Introduction to Law and Legal Reasoning*. Boston: Little, Brown and Co.
—— 1995: *An Introduction to Law and Legal Reasoning*. Boston: Little, Brown and Co., 2nd edn.
Coleman, J. 1982: Negative and positive positivism. *The Journal of Legal Studies*, 11, 139–64.
—— 1995: Truth and objectivity in law. *Legal Theory*, 1, 33–68.
Dworkin, R. 1978: *Taking Rights Seriously*. Cambridge: Harvard University Press.
—— 1985: *A Matter of Principle*. Cambridge: Harvard University Press.
—— 1986: *Law's Empire*. Cambridge: Harvard University Press.
Eisenberg, M. A. 1988: *The Nature of the Common Law*. Cambridge: Harvard University Press.
Fumerton, R. 1994: The incoherence of coherence theories. *Journal of Philosophical Research*, 19, 89–102.
Hobbes, T. [1651] 1962: *Leviathan*, ed. M. Oakeshott, New York: Collier.
Hurley, S. 1990: Coherence, hypothetical cases, and precedent. *Oxford Journal of Legal Studies*, 10, 221–51.
Kennedy, D. 1976: Form and substance in private law adjudication. *Harvard Law Review*, 89, 1,685–778.
Kress, K. 1984: Legal reasoning and coherence theories: Dworkin's rights thesis, retroactivity, and the linear order of decisions. *California Law Review*, 72, 369–402.
—— 1993: Coherence and formalism. *Harvard Journal of Law and Public Policy*, 16, 639–82.
Perry, S. 1995: Interpretation and legal theory. In A. Marmor (ed.), *Law and Interpretation*, Oxford: Clarendon Press, 97–135.
Rawls, J. 1957: Outline of a decision procedure for ethics. *Philosophical Review*, 66, 177–97.
—— 1971: *A Theory of Justice*. Cambridge: Harvard University Press.
Raz, J. 1979: *The Authority of Law*. Oxford: Clarendon Press.
—— 1985: Authority, law and morality. *The Monist*, 68, 295–324.
—— 1992: The relevance of coherence. *Boston University Law Review*, 72, 273–321.
Sartorius, R. 1975: *Individual Conduct and Social Norms: a utilitarian account of social union and the rule of law*. Encino, Ca. Dickinson.

Sayre-McCord, G. 1985: Coherence and models for moral theorizing. *Pacific Philosophical Quarterly*, 66, 170–90.

Walker, R. 1989: *The Coherence Theory of Truth: realism, anti-realism, idealism*. London: Routledge.

Weinrib, E. 1988: Legal formalism: on the immanent rationality of law. *Yale Law Journal*, 97, 949–1,016.

—— 1995: *The Idea of Private Law*. Cambridge: Harvard University Press.

39

The welfare state

SANFORD LEVINSON

What are the proper functions of the state? One answer is the facilitation of decisions made by autonomous individuals, coupled with prevention of the use of force or fraud by insufficiently socialized individuals and provision of a "common defense" against foreign enemies. This answer underlay the classical–liberal, nineteenth-century theory of the "night-watchman state," limited to enforcing private contracts, providing protection against those who violated basic legal norms, and defending the society from hostile incursions. Contemporary theorists of the minimal state would presumably include within the state's proper ambit the provision of certain "public goods," a special set of goods – the usual examples are national defense or the building of dams to prevent flooding – whose enjoyment cannot be limited only to those specific individuals who wish to purchase them. Instead, precisely because there is no effective way to prevent non-purchasers from enjoying what they did not pay for, they become "free riders." As a consequence, most economists would argue, there is underinvestment in the goods in question because of the reluctance of investors to subsidize the free riders. The answer to this problem is to force potential free riders to pay their "fair share" through compulsory taxation. For most economists, though, the category of true "public goods" is relatively restricted, and the state's domain can remain quite limited. Moreover, the taxation and subsequent spending on public goods is in no way redistributive, since by definition the potential "free riders" are compensated for their taxes by the supply of what is stipulated to be a valuable good.

Such views about the minimal role of the state are historically linked with the intellectual development of free-market economics and the economic rise of capitalism. As Gilbert (1983, pp. 4–5) has written, "Capitalism encourages competition and risk-taking behavior," with victory going to those who are the beneficiaries of both their talents and the sheer luck of market vagaries. In turn, however, "misfortune and failure can lead to harsh consequences," for "[t]here are few market mechanisms to mitigate the consequences of accident, illness, ageing, and vicissitudes of industrial society."

As a matter of empirical fact, *no* state has ever adopted a completely minimalist role, however strong its advocacy by such proponents of *laissez-faire* as Herbert Spencer. Whatever the power of that vision, reflected today in the thought of economists like Milton Friedman or the philosopher Robert Nozick (not to mention

A

Court of Appeal

Inland Revenue Commissioners *v* John Lewis Properties plc

[2002] EWCA Civ 1869

2002 Oct 17, 18; Dec 20 — Schiemann, Arden and Dyson LJJ

B

Revenue — Corporation tax — Profits, computation of — Rent factoring agreement — Properties owned by holding company let to trading company — Holding company assigning rental rights to bank in consideration for lump sum payment — Whether payment capital or income receipt

C The taxpayer company was the property holding company for a group of companies involved in the retail trade. It owned the freehold or long leasehold interests in five substantial properties let to the group's trading company. In November 1995 the taxpayer company embarked on a rent factoring scheme for the purposes of obtaining relief from liability for corporation tax. The scheme involved the taxpayer company entering into an agreement with a bank whereby it assigned to the bank for a period of five years the right to receive the rents from the properties D that were payable by the trading company from January 1996 to January 2001. In return the bank agreed to make a lump sum payment of £25·5m to the taxpayer company, a sum calculated to represent a discounted value of the rent. The taxpayer company gave notice of the rental assignment to the trading company, which agreed thereafter to pay the rents as they fell due directly to the bank. By a further agreement the trading company gave warranties and guarantees to the bank that the taxpayer company would perform its obligations under the deed of assignment. The taxpayer company was assessed to corporation tax for its accounting period ending 31 January E 1996 on the basis that the lump sum received from the bank under the deed of assignment was chargeable to corporation tax as either a receipt arising from its ownership of an estate or interest in the properties, under Schedule A of the Income and Corporation Taxes Act 1988, or as an annual profit or gain of an income nature, under Schedule D, Case VI. A special commissioner allowed an appeal by the taxpayer company against the assessment, holding that by virtue of the rental assignment agreement and the payment from the bank of the lump sum the scheme F was effective to relieve the taxpayer company of liability for the tax it would have had to pay on the rents, in place of which it had obtained a capital receipt, that for the purposes of corporation tax the taxpayer company had made part disposals of the properties within section 21(2) of the Taxation of Chargeable Gains Act 1992 and that the taxpayer company and the trading company were entitled to claim roll-over relief on the basis that the assets partly disposed of were buildings occupied and used for the purposes of the trading company's trade. The judge dismissed an appeal by G the Crown.

On appeal by the Crown solely on the issue as to whether the payment was a capital and not an income receipt—

Held, dismissing the appeal (Arden LJ dissenting), that whether a payment was to be regarded as capital or income depended on the nature of the transaction and the factual matrix in which it was set; that the relevant factors in a case such as the present included the duration of the asset disposed of, its value, the extent of any H diminution in the assignor's interest, whether the payment was a single lump sum or a series of recurring payments and whether the disposal was accompanied by a transfer of risk in relation to it; that since the payment by the bank to the taxpayer company was of a single and substantial lump sum which was made in return for the assignment by the company of its rights to receive six years' rent for its properties and which thus had the effect of diminishing the value of its reversionary interests in

them, and since the risk of non-payment of rent was transferred by the taxpayer company to the bank, the payment was to be treated as one of capital and not of income (post, paras 80–87, 94, 99–104, 108–114).

Dicta of Lord Reid, Lord Morris of Borth-y-Gest, Lord Upjohn and Lord Wilberforce in *Strick v Regent Oil Co Ltd* [1966] AC 295, 313, 328, 343, 345, 348, HL(E) applied.

MacNiven v Westmoreland Investments Ltd [2003] 1 AC 311, HL(E) considered.

Paget v Inland Revenue Comrs [1938] 2 KB 25, CA not followed.

Per Arden and Dyson LJJ. Had the receipts been characterised as income, they would have fallen within the charge to tax under either Schedule A or Case VI of Schedule D in the 1988 Act (post, paras 65–66, 104).

Decision of Lightman J [2002] 1 WLR 35 affirmed.

The following cases are referred to in the judgments:

Attorney General v Black (1871) LR 6 Exch 308
British Insulated and Helsby Cables Ltd v Atherton [1926] AC 205, HL(E)
Charterhouse Investment Trust Ltd v Tempest Diesels Ltd [1986] BCLC 1
Comr of Inland Revenue v Wattie [1999] 1 WLR 873, PC
Comr of Taxation of the Commonwealth of Australia v Myer Emporium Ltd (1987) 163 CLR 199
Deeny v Gooda Walker Ltd (No 2) [1996] 1 WLR 426; [1996] 1 All ER 933, HL(E)
Glenboig Union Fireclay Co Ltd v Inland Revenue Comrs 1922 SC(HL) 112, HL(Sc)
Gliksten (J) & Son Ltd v Green [1929] AC 381, HL(E)
Greyhound Racing Association (Liverpool) Ltd v Cooper [1936] 2 All ER 742
Haig's (Earl) Trustees v Inland Revenue Comrs 1939 SC 676
Hallstroms Pty Ltd v Federal Comr of Taxation (1946) 72 CLR 634
Inland Revenue Comrs v British Salmson Aero Engines Ltd [1938] 2 KB 482; [1938] 3 All ER 283; 22 TC 29, CA
Inland Revenue Comrs v McGuckian [1997] 1 WLR 991; [1997] 3 All ER 817, HL(NI)
Inland Revenue Comrs v Wesleyan and General Assurance Society [1946] 2 All ER 749, CA
Jones v Leeming [1930] AC 415, HL(E)
Jones (Henry) (IXL) Ltd v Federal Comr of Taxation (1991) 102 ALR 1
London and Thames Haven Oil Wharves Ltd v Attwooll [1967] Ch 772; [1967] 2 WLR 743; [1967] 2 All ER 124, CA
Lowe v J W Ashmore Ltd [1971] Ch 545; [1970] 3 WLR 998; [1971] 1 All ER 1057
McClure v Petre [1988] 1 WLR 1386
MacNiven v Westmoreland Investments Ltd [2001] UKHL 6; [2003] 1 AC 311; [2001] 2 WLR 377; [2001] 1 All ER 865, HL(E)
Paget v Inland Revenue Comrs [1938] 2 KB 25; [1938] 1 All ER 392, CA
Raja's Commercial College v Gian Singh & Co Ltd [1977] AC 312; [1976] 3 WLR 58; [1976] 2 All ER 801, PC
Ramsay (W T) Ltd v Inland Revenue Comrs [1982] AC 300; [1981] 2 WLR 449; [1981] 1 All ER 865, HL(E)
Strick v Regent Oil Co Ltd [1966] AC 295; [1965] 3 WLR 636; [1965] 3 All ER 174; 43 TC 1, HL(E)
Sun Newspapers Ltd v Federal Comr of Taxation (1938) 61 CLR 337
Vallambrosa Rubber Co Ltd v Farmer 1910 SC 519
Van den Berghs Ltd v Clark [1935] AC 431, HL(E)
West Midland Baptist (Trust) Association (Inc) v Birmingham Corpn [1970] AC 874; [1969] 3 WLR 389; [1969] 3 All ER 172, HL(E)
Withers v Nethersole [1946] 1 All ER 711, CA; [1948] 1 All ER 400, HL(E)

The following additional cases were cited in argument:

British and Commonwealth Holdings plc v Barclays Bank plc [1996] 1 WLR 1; [1996] 1 All ER 381, CA

A *Furniss v Dawson* [1984] AC 474; [1984] 2 WLR 226; [1984] 1 All ER 530, HL(E)
Sothern-Smith v Clancy [1941] 1 KB 276; [1941] 1 All ER 111, CA

APPEAL from Lightman J

The taxpayer company, John Lewis Properties plc ("JLP"), the property holding company for the John Lewis Partnership group of companies, was assessed to corporation tax for the accounting period to 31 January 1996 in B an amount of £31m on the basis that a sum of £25·5m paid to it by a Dutch bank, Cooperatieve Centrale Raiffeisen-Boerenleenbank BA ("Rabobank"), under the terms of a rental assignment agreement dated 20 November 1995, was an income and not a capital receipt. On 5 September 2000 a special commissioner, Mr T H K Everett, allowed the taxpayer's appeal against that assessment. On 13 June 2001 the judge dismissed an appeal by the Crown C under the provisions of section 56A of the Taxes Management Act 1970.

The Crown appealed by permission of the judge on the following grounds. (1) The judge erred in law in holding that the payment was a capital receipt in the hands of JLP. He ought to have held that it was an income receipt. (2) Having correctly identified the basic principles of law by reference to which the question was to be determined and having correctly held that he was not bound by the Court of Appeal decision in *Paget v Inland* D *Revenue Comrs* [1938] 2 KB 25 to hold that it was a capital receipt, the judge ought to have held the payment to be an income receipt, because (a) it was obtained for the short-term disposal of income-producing assets that were retained by JLP throughout in its capacity as reversioner, and (b) from a practical and business point of view the receipt was calculated not to effect an outright disposal of the properties but rather the conversion of future E income into present income. (3) Strong support for the Crown's contention was provided by the Australian authorities, in particular *Comr of Taxation of the Commonwealth of Australia v Myer Emporium Ltd* (1987) 163 CLR 199 and *Henry Jones (IXL) Ltd v Federal Comr of Taxation* (1991) 102 ALR 1. (4) There was no English or United Kingdom case precluding the decision that the receipt was income: it contradicted the dictum of Lord F Romer in *Paget's* case but that dictum did not form part of the ratio, was wrong in law and could not stand with the House of Lords' reasoning in *MacNiven v Westmoreland Investments Ltd* [2003] 1 AC 311. (5) Had JLP been deprived of six years' rent from the properties by the act of a tortfeasor, any damages recovered by JLP as compensation would have been taxable as an income receipt. (6) The judge erred in holding that Lord G Hoffmann's guidance in *MacNiven's* case supported the approach of Lord Romer in *Paget's* case and/or reinforced the view that the receipt was a capital one. From a commercial point of view, the receipt was income and did not represent an exchange of income for capital. The Crown's case was not based on economic equivalence but on the proper characterisation of the receipt in the light of the principles derived from the authorities. (7) The judge erred in deriving support for his conclusion from (a) the premise on H which *Inland Revenue Comrs v McGuckian* [1997] 1 WLR 991 was originally decided by the House of Lords and was reviewed in *MacNiven's* case, (b) the proposition that a premium paid for a lease was capital and (c) the provisions of the Finance Act 2000 enacted to counter rent factoring schemes.

By a respondent's notice dated 15 August 2001 the taxpayer company contended that if the judge had erred in concluding that the receipt was one of capital and if it was to be properly characterised as income, it would not be chargeable to tax as income because it would not fall within (a) the wording of Schedule A, as set out in section 15 of the Income and Corporation Taxes Act 1988, (b) the wording of Case VI of Schedule D, as set out in section 18 of the 1988 Act, or (c) any other Schedule or Case of the 1988 Act, so that there would be no basis on which it could be charged to tax as income under the schedular system of tax in force in the United Kingdom.

The facts are set out in the judgment of Arden LJ.

Launcelot Henderson QC and *Michael Furness QC* for the Crown. The issue is whether the price received by JLP for the assignment of the rents to the bank was an income or a capital receipt in the hands of JLP. Principles of English law determine that a receipt for the disposal of the recurrent produce of an asset is income: see *Lowe v J W Ashmore Ltd* [1971] Ch 545 and *McClure v Petre* [1988] 1 WLR 1386. The character of the receipt does not depend on the juristic classification of the transaction but rather on a practical, business analysis of what it effected.

Lord Romer's dictum in *Paget v Inland Revenue Comrs* [1938] 2 KB 25, 44–46, to the effect that the proceeds of sale of a right to receive income in the future cannot be treated as income for tax purposes is incompatible with the well-established principles laid down in *Comr of Inland Revenue v Wattie* [1999] 1 WLR 873 and *Hallstroms Pty Ltd v Federal Comr of Taxation* (1946) 72 CLR 634. Further, it was incompatible with Lord Hoffmann's reasoning in *MacNiven v Westmoreland Investments Ltd* [2003] 1 AC 311, 326–336. The judge erred in holding that the decisions in *MacNiven's* case and *Inland Revenue Comrs v McGuckian* [1997] 1 WLR 991 lead to the conclusion that the principle enunciated by Lord Romer compelled characterisation of the price as capital. *MacNiven's* case establishes that where a right to income is assigned in return for a payment of money the payment will be a capital receipt in the hands of the assignor if from a commercial point of view the transaction amounts to an exchange of income for capital, eg, the purchase/sale of an annuity: see *Sothern-Smith v Clancy* [1941] 1 KB 276, 283. *McGuckians's* case decided that a sale of the right to a dividend should be disregarded for fiscal purposes, in accordance with what was then understood to be the classic statement of the anti-avoidance principle in *W T Ramsay Ltd v Inland Revenue Comrs* [1982] AC 300, as stated by Lord Brightman in *Furniss v Dawson* [1984] AC 474, 527. *MacNiven's* case establishes that that approach was wrong, but that the sale nevertheless failed to transform the receipt of a dividend into a capital sum. There cannot be a principle of tax law that an assignment of a right to future income necessarily produces a capital receipt.

The price was income because it constituted (a) a receipt for the short-term disposal of the inherently recurrent produce of income-producing assets retained by JLP throughout in its capacity as reversioner, and (b) a receipt calculated to effect from a practical and business point of view, not the outright disposal of the properties or any part of the properties, but rather the conversion of future income (rents) into present income (the price). The price was calculated as the discounted value of the assigned rents. JLP retained full beneficial ownership of the properties. After the

A expiry of the term of the assignment JLP would be entitled, as before, to the rental income. The transaction did not as a matter of business reality transfer the risk of non-payment of rents from JLP to the bank. What matters is the business character of the receipt in JLP's hands. The Crown does not ask the court to recharacterise a capital payment but simply to characterise the price as income, it not being paid for the acquisition of an B enduring asset in any business sense. Strong support for the Crown's approach is provided by the decision of the High Court of Australia in *Comr of Taxation of the Commonwealth of Australia v Myer Emporium Ltd* (1987) 163 CLR 199, as explained and applied in *Henry Jones (IXL) Ltd v Federal Comr of Taxation* (1991) 102 ALR 1. Further support is provided by cases concerning compensation which establish that a payment that fills a hole in a taxpayer's income receipts, or compensates him for liability to C make income payments, is itself an income receipt: see *Raja's Commercial College v Gian Singh & Co Ltd* [1977] AC 312. [Reference was also made to *Deeny v Gooda Walker Ltd (No 2)* [1996] 1 WLR 426; *London and Thames Haven Oil Wharves Ltd v Attwooll* [1967] Ch 772; and *J Gliksten & Son Ltd v Green* [1929] AC 381.] Similarly, a prepayment of rent would itself have been an income receipt: see *Greyhound Racing Association (Liverpool)* D *Ltd v Cooper* [1936] 2 All ER 742.

David Goldberg QC and *Wayne Clark* for the taxpayers. The price is capital. It is a fundamental principle of income tax law that the realisation of the value of a capital asset produces a capital and not an income receipt. Nothing turns on the legal form of the transaction. The issue is a commercial one: did the relevant transaction realise value from a capital E asset? If the answer is "yes" the receipt is and must be capital because there is no basis on which it can be income: see *McClure v Petre* [1988] 1 WLR 1386, 1393. The sale of the rents is a part disposal of the properties and effects a realisation of part of their value with exactly the same commercial and economic effects as other similar transactions of sale or lease would have done.

Whether a receipt is income or not falls to be answered by reference to F what may be called the common law of tax rather than by reference to statutory provisions. The courts' approach is to consider the receipt from a practical and business point of view rather than by reference to the "juristic classification of the legal rights": see *Comr of Inland Revenue v Wattie* [1999] 1 WLR 873, 880 and *Hallstroms Pty Ltd v Federal Comr of Taxation* (1946) 72 CLR 634, 648. Where a receipt is from a transaction which G realises the value of a capital asset, the receipt is capital: see *Paget v Inland Revenue Comrs* [1938] 2 KB 25, 35, 44–45. [Reference was also made to *Inland Revenue Comrs v McGuckian* [1997] 1 WLR 991, 996; *Comr of Taxation v Wattie* [1999] 1 WLR 873 and *MacNiven v Westmoreland Investments Ltd* [2003] 1 AC 311, 332–334.] The form of the transaction is irrelevant: see *Glenboig Union Fireclay Co Ltd v Inland Revenue Comrs* 1922 SC(HL) 112, 114; *Earl Haig's Trustees v Inland Revenue Comrs* 1939 H SC 676, 681–683; *Withers v Nethersole* [1948] 1 All ER 400; *McClure v Petre* [1988] 1 WLR 1386, 1389–1391 and *Strick v Regent Oil Ltd* [1966] AC 295, 341–342, 348–349.

The assignment of the rents had commercial reality: it reduced the value of the reversion to the properties and involved a transfer of risk, the risk that

JL would not pay the rents being transferred from JLP to the bank. The price paid by the bank was not the rents relabelled but a distinct payment made before the rents were even payable. The price is capital in JLP's hands from a practical and a business point of view, being derived from a part disposal of the properties and indistinguishable from any other sale proceeds arising from them. Section 110 of the Finance Act 2000 has the effect of reinforcing the capital character of the rental assignment. The issues are determined by the correct and binding decisions in *Paget v Inland Revenue Comrs* [1938] 2 KB 25 and *McClure v Petre* [1988] 1 WLR 1386 and thus no charges to corporation tax on income can arise in respect of the receipts.

However, if the price does constitute income, it does not fall within any of the Schedules or Cases set out in the Income and Corporation Taxes Act 1988. The judge erred in concluding that it is within Schedule A (set out in section 15 of the Act) because the price is not a receipt arising from or by virtue of the ownership of land. *Lowe v J W Ashmore Ltd* [1971] Ch 545 does not support the judges's conclusion. Nor is the price, being payable on a transaction of sale and purchase, akin to other income described in Schedule D within Case VI of that Schedule: see *Attorney General v Black* (1871) LR 6 Exch 308; *Jones v Leeming* [1930] AC 415 and *Withers v Nethersole* [1948] 1 All ER 400.

Henderson QC replied.

Cur adv vult

20 December. The following judgments were handed down.

ARDEN LJ

1 The Inland Revenue appeals against the order dated 13 June 2001 of Lightman J [2002] 1 WLR 35 dismissing an appeal by it against the decision [2000] STC (SCD) 494 of a special commissioner, Mr T H K Everett, dated 5 September 2000, on the grounds that the sum of £25,556,762·55 (which I will call "the proceeds") received by the taxpayer, John Lewis Properties plc ("JLP"), as consideration for an assignment of rents constituted a capital, and not an income, receipt. Receipts are chargeable to corporation tax whether they are income or capital, but only if they are capital does the taxpayer have the ability to claim certain reliefs including the ability to roll over any gain into the costs of improving the properties.

2 There is no dispute about the facts. In essence, JLP owns the freehold or a long leasehold interest in five substantial properties which are let to John Lewis plc ("JL") on yearly tenancies. JL is the ultimate group holding company. In 1995 JLP entered into arrangements for factoring some of the rents receivable from the properties. By a deed of assignment dated 20 November 1995 made between JLP and a Dutch bank, Cooperatieve Centrale Raiffeisen-Boerenleenbank BA ("Rabobank"), JLP assigned to Rabobank the right to receive rents payable in respect of the five years and one day between 23 January 1996 up to and including 23 January 2001 in return for the proceeds. The proceeds were calculated as the value of the rents due to be received in that period, using a discount rate of about 7·6%. (There is no suggestion that that rate was not negotiated at arm's length.) JLP gave notice of the assignment to JL. Rabobank and JLP also entered into a swap arrangement pursuant to which JLP would pay to or receive from

A Rabobank the amount by which a notional commercial floating rate of interest on an amount equal to the consideration for the assignment exceeded, or was less than, the discount rate of 7·6%. By a further agreement of the same date made between JLP, JL and Rabobank, JL gave certain warranties and undertakings to Rabobank in relation to its and JLP's financial position and, among other matters, agreed to indemnify Rabobank B against non-payment of the assigned rents. Moreover, under clause 16 of the guarantee and indemnity Rabobank could require the transaction to be unwound if JL failed to pay any of the assigned rents.

3 Following the rental assignment, the bulk of the proceeds were used by the JL group to open new stores and to improve its existing stores. JLP now claims rollover relief in respect of this expenditure. In accordance with Financial Reporting Standard 5, Reporting the Substance of Transactions, C JLP has prepared its statutory accounts (which are required to show a true and fair view) by accounting for the proceeds as an advance in its accounts and the assigned rents for the period under review as receivables, because the risk of non-payment remains with JLP. However, little argument has been addressed to Financial Reporting Standard 5 on this appeal. It is not clear to me why that course was taken since the question whether a receipt is capital D or income has to be decided from a commercial point of view and in principle the accountancy treatment is, therefore, a relevant consideration.

4 There was no evidence from any officer or employee of JLP as to why the assignment was executed. There was expert evidence from a chartered accountant, Mr Philip Haberman, who analysed the theoretical distinctions between the position of JLP under the transaction and the position of JLP on E the basis that no transaction had been carried out. These distinctions included a reduction in the value of the underlying properties, certainty of receipt, certainty as to timing, removal of economic risk, removal of regulatory risk, creation of opportunity and removal of administrative obligations, but it is not suggested that these distinctions were the actual reasons why JLP carried out the transactions. Mr Haberman also produced what he described as practical differences in economic terms. These F included a table showing "the opportunity value" to the John Lewis Partnership group of the assignment on a pre-tax basis. He carried out the same exercise on a post-tax basis. Using the group's actual return on capital, he was able to show that the group's return on capital is greater if it received a sum of £25m in year I than if it received the rents due each year. In cross-examination Mr Haberman agreed that the assignment was a financing G operation and could be viewed as either a loan or purchase. The special commissioner [2001] STC 1118, 1124 accepted Mr Haberman's evidence. JLP called a further expert witness, Mr Richard Asher, a chartered surveyor. The opinion of Mr Asher was that:

> "The value of JLP's reversionary interests in the properties would be H reduced immediately following the rental assignment as any purchaser of those interests would have to take subject to the rental assignment and would only acquire the right to receive rentals after the five-year rental assignment period had expired. The value of the properties should gradually increase again as the period of the rental assignment outstanding reduces over time."

The special commissioner, at p 1124, also accepted the evidence of Mr Asher. However, the amount of the reduction in value which in Mr Asher's opinion would occur was not quantified by him as at any point in time.

5 On this appeal it is common ground that the deed of assignment effected both an assignment of the contractual rights of JLP to receive the rents and the transfer of an interest in land. The judge was required to spend considerable time on that issue although it is of limited significance in relation to the capital/income issue for the reasons given below.

6 The only questions decided by the judge which arise on this appeal are the questions whether the proceeds constituted a capital or income receipt, and if the latter under which head of charge they were chargeable. The judge analysed the capital/income issue in these terms [2002] 1 WLR 35, 43, para 18:

> "The authorities offer as a guide the principle that a receipt for the recurrent produce of an asset and compensation for the loss of such produce or to make good a hole in the receipts from such produce constitute income, whilst a receipt for the asset or part of the asset or for the permanent impairment or the sterilisation of an asset constitutes capital."

In so doing the judge drew the hallowed distinction between the fruit of a tree and the tree itself. He further concluded from his survey of the authorities that it was clear that the receipt of a lump sum in consideration of the sale of an income stream, together with the underlying asset producing such income stream, was a capital receipt, as was the receipt of a lump sum in consideration of an income stream when there was no underlying asset (for example, an annuity). In his judgment, the question raised by the present case was whether the position was the same where a lump sum was paid in consideration for an income stream but the underlying asset was retained by the vendor.

7 On this, the judge referred to *Paget v Inland Revenue Comrs* [1938] 2 KB 25, 44–45 in which Lord Romer, sitting in the Court of Appeal, held that the proceeds of the sale for a lump sum of an annuity were capital and that this was so even where the subject of the sale was not the annuity for its whole duration but the right to be paid the annuity for a number of years or even for one. On this appeal, as before the judge, the revenue challenges the correctness of Lord Romer's conclusions in this respect.

8 The judge then examined certain Australian cases which in the event he did not find persuasive as they conflicted with what Lord Romer had said in *Paget's* case. The judge also examined two recent decisions of the House of Lords, namely *Inland Revenue Comrs v McGuckian* [1997] 1 WLR 991 and *MacNiven v Westmoreland Investments Ltd* [2003] 1 AC 311. In the latter case the House of Lords restated the principles established in *W T Ramsay Ltd v Inland Revenue Comrs* [1982] AC 300. I shall have to examine these cases below but, critically, Lord Hoffmann indicated in *MacNiven's* case [2003] 1 AC 311, 333, para 54 that a transaction without commercial reality which purported to exchange income for capital failed "to perform the alchemy of transforming the receipt of a dividend from the company into the receipt of a capital sum from someone else".

A 9 Having analysed the speech of Lord Hoffmann, with which the remainder of the House agreed, the judge expressed his conclusions thus [2002] 1 WLR 35, 50, para 37:

> "The guidance afforded by Lord Hoffmann in my view supports the approach of Lord Romer in *Paget v Inland Revenue Comrs* [1938] 2 KB 25 and reinforces the view that the price received by JLP was capital and B not income: (i) JLP was perfectly entitled for the avoidance of tax to structure its commercial transaction with the bank so that in place of an income receipt of rent it received a capital sum. There is no broad 'economic equivalence test' entitling the court to treat a capital item as income because it is the economic equivalent of income; (ii) the transaction produced in the hands of JLP, in place of an income stream, an C up-front capital sum; (iii) the [proceeds were] not merely (as in *McGuckian's* case [1997] 1 WLR 991) the bank's receipt of the rents from the lessees 'relabelled': it was a distinct sum paid out of the resources of the bank under a transaction which had commercial reality; (iv) in these circumstances it is not open to the court to recharacterise the [proceeds] as income; (v) the reference by Lord Hoffmann to exchanges of income for capital involving a transfer of risk does not mean that there D can be no such exchange for tax purposes in any particular case unless there arises a substantial risk of loss through non-payment which is unsecured or unprovided for. In any event there is in this case a theoretical risk of loss undertaken by the bank if JLPG, JLP and JL went into insolvent liquidation."

10 The judge held that his conclusion was further supported by the E premise on which the *McGuckian* case was decided, namely that the exchange of an income stream for a lump sum prima facie constituted capital: see the speech of Lord Browne-Wilkinson [1997] 1 WLR 991, 995–996. However, the judge recognised that he was not bound by a proposition assumed to be correct by the House of Lords but this reinforced his view.

F 11 The judge rejected the argument that amendments made by the Finance Act 2000 (to insert into the Income and Corporation Taxes Act 1988 provisions treating the proceeds of certain rent factoring agreements as rent) shed much light on the law in force in 1995. He also rejected the argument that the proceeds of the deed of assignment, if taxable as income, were not within any head of charge.

12 The contention for the revenue is that the proceeds of the deed of G assignment constitute income because they were obtained for the short-term disposal of inherently recurrent produce of income-producing assets which were retained by JLP and because the receipt was calculated to effect from a business and practical point of view not the outright disposal of the properties or any part thereof but rather the conversion of future income into present income. The contention of the taxpayer is that the deed of H assignment effected the sale of part of the asset constituted by each property, alternatively part of the value of a capital asset, and accordingly the proceeds represent a capital receipt.

13 The approach which the court must adopt is one of practical and business reality. Thus, Dixon J in *Hallstroms Pty Ltd v Federal Comr of Taxation* (1946) 72 CLR 634, 648, in a passage cited by Lord Nolan when

giving the advice of the Privy Council in *Comr of Inland Revenue v Wattie* [1999] 1 WLR 873, 880, held that the answer to the question whether expenditure is of a capital or revenue nature "depends on what the expenditure is calculated to effect from a practical and business point of view, rather than upon the juristic classification of the legal rights, if any, secured, employed or exhausted in the process".

14 Sir Nicolas Browne-Wilkinson V-C summed up the approach in a similar way in *McClure v Petre* [1988] 1 WLR 1386, 1389:

> "In my judgment it is equally established by authority that to decide whether a particular receipt is in the nature of income or in the nature of capital one has to look at all the circumstances of the particular case and apply judicial common sense in reaching a conclusion as to how a receipt is to be classified."

15 There is further valuable guidance from the House of Lords in *MacNiven's* case [2003] 1 AC 311. I refer to this below.

16 In the context of capital and income (more precisely, in the context of converting income into capital), the House of Lords has used the word "alchemy", and it has to be observed that this court has not yet found the philosopher's stone, if one exists. One eminent Master of the Rolls, Lord Greene, expressed the "somewhat cynical" view (per Lord Upjohn in *Strick v Regent Oil Co Ltd* [1966] AC 295, 345) that "in many cases it is almost true to say that the spin of a coin would decide [the question whether a receipt was income or capital] almost as satisfactorily as an attempt to find reasons": *Inland Revenue Comrs v British Salmson Aero Engines Ltd* [1938] 2 KB 482, 498. A number of cases have been cited on this appeal on either side of the line. It is difficult at first sight to discern any organising principle save in the "compensation" cases (see below), where compensation is treated as capital or income according to the nature of the sum being replaced, and (perhaps) the case of premiums on leases. However, the decided cases certainly shed light on the qualities of capital and income. I take the most significant of them in turn. In grouping them I am immediately faced by the usual snares of taxonomy. I propose to group the English and Scottish cases cited to us as follows.

(1) Cases illustrating the realisation principle: *Glenboig Union Fireclay Co Ltd v Inland Revenue Comrs* 1922 SC(HL) 112, *Paget's* case [1938] 2 KB 25, per Lord Romer, *Earl Haig's Trustees v Inland Revenue Comrs* 1939 SC 676, *Withers v Nethersole* [1946] 1 All ER 711; [1948] 1 All ER 400, *Lowe v J W Ashmore Ltd* [1971] Ch 545 and *McClure v Petre* [1988] 1 WLR 1386. I use the expression "realisation principle" to denote the principle decided by these cases. "Sterilisation" might have been a better word.

(2) Cases on premiums on leases: *Strick v Regent Oil Co Ltd* [1966] AC 295 and *Wattie's* case [1999] 1 WLR 873.

(3) Compensation cases: the *Glenboig* case 1922 SC(HL) 112, *J Gliksten & Son Ltd v Green* [1929] AC 381, *London and Thames Haven Oil Wharves Ltd v Attwooll* [1967] Ch 772, *Raja's Commercial College v Gian Singh & Co Ltd* [1977] AC 312 and *Deeny v Gooda Walker Ltd (No 2)* [1996] 1 WLR 426.

(4) Cases on the sale of rights to income: *Paget's* case [1938] 2KB 25, the *McGuckian* case [1977] 1 WLR 991 and *MacNiven's* case [2003] 1 AC 311.

A 17 As can be seen, at least one case falls into more than one group and the grouping, therefore, is not watertight and only serves the purpose of exposition.

The realisation principle

18 The *Glenboig* case 1922 SC(HL) 112 is a classic illustration of the B realisation principle. The taxpayer company was a merchant of raw fireclay which it extracted from certain fireclay fields which it leased and over which the Caledonian Railway ran. The railway company, no doubt fearful of subsidence, obtained an order preventing it from working these fields and in due course there was a settlement agreement whereby the taxpayer agreed not to work the fireclay fields in return for compensation. The question considered by the House of Lords was whether the compensation, which the C railway company paid to the taxpayer as compensation for not being able to work these fireclay fields, constituted a revenue or capital receipt. The compensation represented two and a half years' profits from the fireclay fields, being the period of the remaining working life of the fields. The House of Lords unanimously held that the compensation was a capital receipt. Lord Buckmaster, at p 115, considered it irrelevant that the D compensation was calculated by reference to expected trading profit: "It is unsound to consider the fact that the measure adopted for the purpose of seeing what the total amount should be was based on considering what were the profits that would have been earned." The important factor was "the quality of the figure that is arrived at by means of the test" of that measure: p 115. In the words of Lord Wrenbury, at p 116, the compensation "was the price paid for sterilising the asset from which otherwise profit might have E been obtained. What is true of the whole must be equally true of part."

19 *Earl Haig's Trustees'* case 1939 SC 676, a decision of the Inner House of the Court of Session, is an illustration of the same principle, but the realisation in that case was of the publication rights of Earl Haig's diaries. The realisation was thus only partial, but the publication rights were considered to have been largely exhausted and they represented a large part F of the asset: pp 682–686. As Lord Fleming put it, the transaction was not one which merely involved a use of the subject salva rei substantia (leaving it substantially intact). Again the amount of the receipt was calculated by reference to the amount of profits received from publication. The revenue on this appeal have adopted the well-worn analogy between a tree and its fruit. It has been used many times before and it was also relied on by Lord G Moncrieff in this case, at p 687:

> "the differentiation in any particular case of profits as resulting in capital appreciation on the one hand or, on the other, in accruer of income may be expected to introduce questions of difficulty. In determining any such question, however, it may perhaps be helpful to distinguish between a fruit-bearing subject and its fruits. The right of H usufruct as conferred by the civil law and so described, or again our own requirement that liferents be enjoyed salva rei substantia, may perhaps be regarded as illustrative. When fruits are ingathered and realised, this in general results in an addition to income; and this will equally be true in the case of natural fruits, industrial fruits, or civil fruits. When the subject which provides the fruits is realised on the other hand, this generally

results in a transaction on capital account. He must expect a loss, and not a gain, of income, who kills the goose that lays the golden eggs." A

20 In *Nethersole's* case [1946] 1 All ER 711 a dramatist, Miss Nethersole, who had retired, received payments under an agreement with the author (Rudyard Kipling) of a work (*The Light that Failed*) that she had dramatised. Pursuant to the agreement, the executrix of the author disposed of the film rights (only) in the book and play for ten years in exchange for a lump sum and the payment in issue represented Miss Nethersole's share of the payment the executrix received on that disposal. This agreement operated as a partial realisation of her copyright in the play. The Court of Appeal held that that sum was capital in Miss Nethersole's hands. Lord Greene MR, giving the judgment of the court, rejected the idea that there could be a principle that the disposal for consideration of limited rights in property, so that the property on expiration of the rights came back to the owner intact, should always be treated as resulting in a revenue receipt whereas if the value of the property was permanently diminished by the disposal there was clearly a disposal of capital: pp 715–716. He referred to the payment of a lump sum in consideration of the grant of a patent licence (in addition to royalties) in the *British Salmson Aero Engines Ltd* case [1938] 2 KB 482, where such a payment was held (by a court also including Lord Greene MR) to be capital. He held, however [1946] 1 All ER 711, 716: B C D

"If the lump sum is arrived at by reference to some anticipated quantum of user it will, we think, normally be income in the hands of the recipient. If it is not, and if there is nothing else in the case which points to an income character, it must, in our opinion, be regarded as capital." E

In *Nethersole's* case the lump sum was not related to any royalty calculation. The Court of Appeal concluded that the lump sum was a capital receipt. Moreover, Miss Nethersole was being paid for, among other things, the right to cut her play to pieces and combine the story with other stories, a right which, whether it should be exercised or not, amounted to a right to diminish the value of the copyright in the play. F

21 The Crown appealed to the House of Lords, who affirmed the decision of the Court of Appeal. Viscount Simon said [1948] 1 All ER 400, 403:

"Here we have the sale and transfer outright of an item of property which previously belonged to the taxpayer, not the licence to use it granted by its unchanged owner, and this does not give rise to annual profits or gains unless the sale takes place in the course of carrying on a trade or profession." G

22 Lord Uthwatt observed, at p 405: H

"The relevant fact is that an owner of an asset, entitled by law to divide it into two distinct assets, has done so by selling one of those assets for an agreed consideration payable in a lump sum. A sale, not in the course of trade, of an asset does not attract tax on the consideration."

A It is to be noted that *Nethersole's* case concerned the sale of part of the copyright to a dramatic work, not the sale of royalties earned under a licence to use the copyright work.

23 In *McClure v Petre* [1988] 1 WLR 1386 an owner of land received a lump sum for a licence to deposit topsoil on his land up to the maximum permitted by law. There could never be a recurrence of such a payment. B There was a once and for all disposal of a valuable right or advantage and the receipt was therefore capital.

24 The observations of Lord Romer in *Paget v Inland Revenue Comrs* [1938] 2 KB 25 also fall within this group. Miss Paget held bearer bonds of the City of Budapest (payable in London) and the kingdom of Yugoslavia (payable in New York) with interest coupons attached. The issuers of the bonds were in default and substitutionary payments were proposed. C Miss Paget sold the interest coupons, which were evidently detachable bearer obligations, in London. The case is an example of the realisation principle because Miss Paget disposed of all her rights in the interest coupons which she sold. The question was whether Miss Paget was liable to income tax on the sale proceeds under either Schedule C or Case IV of Schedule D. Schedule C was inapplicable and Schedule D applied only if the sale D proceeds were "income arising from securities out of the United Kingdom". The part of the decision relevant to this appeal is the part rejecting the contention that the sale proceeds fell to be taxed under Case IV of Schedule D. Sir Wilfrid Greene MR said, at p 35:

> "[This] contention . . . can be disposed of quite shortly. The purchase price received by Miss Paget was not income arising from the bonds at all. E It arose from contracts of sale and purchase whereby Miss Paget sold whatever right she had to receive such income in the future as well as her right to take what was offered by the defaulting debtors. It is, in my opinion, quite impossible to treat this as equivalent in any sense to 'income arising from' the bonds."

25 The second member of the court, Lord Romer, said, at pp 44–45:

F

> "In these circumstances, the only question to be decided is whether the proceeds of sale of a right to receive income in the future can be treated as income for the purpose of the Income Tax Acts. The question thus broadly stated plainly admits of but one answer; and that answer must be in the negative. The proceeds of the sale for a lump sum of an annuity, for instance, are capital in the hands of the vendor and not income. And this G is true even when the subject of the sale is not the annuity for its whole duration but the right to be paid the annuity for a number of years or even for one year."

26 To MacKinnon LJ the issue was clear: the issuer had paid no interest and the sale proceeds could not be "income from securities": see p 48.

27 All three members of the court therefore rejected the argument that H Miss Paget's sale proceeds constituted income within Schedule D. No cases are cited on the capital/income distinction. While Lord Romer clearly held that the revenue's claim had to be rejected because the sale proceeds were capital, the ratio of the majority was that the receipt did not derive from the bonds in any event. The actual result in *Paget's* case was reversed by statute:

see now section 730 of the Income and Corporation Taxes Act 1988, referred to in paragraph 51 below. A

28 I thus reject the submission of Mr Goldberg, for JLP, that this court is bound by what Lord Romer said and that *Paget's* case concludes this case in the taxpayer's favour. Lord Romer's observations must, however, be given great respect, especially as they have stood unchallenged for nearly 75 years. I have concluded, however, that, read generally as supporting the taxpayer's contention that the sale of the right to receive income produces capital, Lord Romer's observations cannot stand with later cases, for example, *Withers v Nethersole* [1948] 1 All ER 400. These cases make it clear that there is no inflexible rule that the sale of an asset, or part of an asset, for a lump sum will always be treated as resulting in a capital receipt: it will however be capital if there is a disposal of the asset, or part of the asset, which produces income. This is so even if the asset itself only has a limited life. I further accept the submission of Mr Henderson, for the Crown, that an inflexible rule that the sale of an asset always produces a capital receipt would be inconsistent with the required approach of examining all the circumstances from a practical and business point of view. On the other hand, in the context of the case, there is a similarity in point of fact between the sale of a detachable interest coupon and an annuity. As the High Court of Australia (see below) has pointed out, they are both sales of the right to receive a payment of income for which there is no underlying asset. (There would moreover be likely to be a separate coupon for each interest payment due over the life of the bonds in *Paget's* case.) I accept that the sale of a single year's payment of an annuity might fall within the realisation principle even though it is only a partial realisation of the asset since that part of the asset is then disposed of. On the other hand, it would not always do so. In the case where the annuitant is entitled to, say, six monthly payments of annuity I do not consider that he could necessarily avoid income tax by selling his right to a six months' instalment for a lump sum to a third party. I thus distinguish between an annuity (or part of an annuity) and instalments of annuity. I also distinguish between an interest coupon and payments of interest thereon. The annuity and the interest coupon are assets from which income is derived and are entire in themselves. Lord Romer's observations are consistent with the realisation principle when that distinction is kept in mind. Turning to this case, I consider that the right to receive rent is analogous to a payment or instalment of an annuity rather than to a part of the annuity itself. The right to receive rent is a payment for the user of property and is thus dependent on the performance by the landlord or lessor of the covenants contained in the tenancy or lease. The right to receive rent is thus not analogous to an annuity, which is entire in itself. The chose in action is not in reality the asset from which profit in the form of rent is earned. In a practical sense, such a chose in action is simply income—the fruit of the tree or (per Lord Moncrieff) the golden eggs rather than the goose itself. It can likewise be said that the right to receive distributions made by a company is not in reality the asset from which profit in the form of distributions is earned. Distributions are paid out of profits earned on the assets representing a company's share capital and reserves: all these represent the tree while distributions declared (and the assets earmarked to pay them) and the right to receive future distributions represent the fruit. The *McGuckian* case [1997] 1 WLR 991 (considered below) is not, however, authority for the

A proposition that the right to future distributions, if assigned, gives rise to a capital receipt because that proposition was there conceded, and was not the subject of decision by the House of Lords.

29 In *Lowe v J W Ashmore Ltd* [1971] Ch 545 the issue was whether the abstraction of turfs by a farmer and the receipt by him of proceeds of sale resulted in a capital or income profit. There was no evidence as to how long it was before there could be turfs again on the land fit for sale. However, B Megarry J inferred that by proper cultivation of the land it would be possible to produce turfs again, at any rate after some period. Accordingly, Megarry J held, at p 551, that the land remained capable of producing again the products which had been sold and that accordingly the receipts were income. This case, therefore focuses on the element of "recurrence" in the asset and, since the rents payable in respect of the properties have the same C quality, Mr Goldberg has argued persuasively that this case can be distinguished. Mr Goldberg submitted that in reaching this conclusion the court had been influenced by the fact that the taxpayer was trading as a farmer. In my judgment this is not correct, as Megarry J was dealing with the question of whether the profits were "annual profits" within section 122 of the Income Tax Act 1952. It was only when he reached his conclusion on that point that he went on to consider whether the profits arose from a trade. D The "recurrence capability", that is the ability of the asset to produce more assets of the type transferred, has been held to be a relevant test of income. One only needs to recall the analogy of the fruit and the tree to realise that, as a matter of common sense, recurrence must be a sign of income.

30 In passing Mr Henderson drew the court's attention to the fact that in *McClure's* case [1988] 1 WLR 1386, 1392 Sir Nicolas Browne-E Wilkinson V-C said that the proposition to be derived from *Lowe's* case is "obviously also limited to the case where there has been no disposal of the land or any interest in it". He submitted that in *Lowe's* case [1971] Ch 545, 557–558 Megarry J clearly contemplated that the right to take turf could properly be classified as a profit à prendre in gross, although he did not decide the point. Be that as it may, in the present context it is of little moment whether JLP technically disposed of an interest in land by the F assignment because the character of the receipt as income or capital does not depend on its juristic classification.

31 Accordingly, all the cases in this group, other than *Paget v Inland Revenue Comrs* [1938] 2 KB 25 and *Lowe v J W Ashmore Ltd* [1971] Ch 545, concern the sale of the whole or part of an asset from which profits may be earned. I have explained that, in my view, the right to receive rents is not G an asset of this kind. I would also observe that it was also apparently not relevant in *Lowe's* case to inquire whether the value of the farm had in fact been diminished by the sale of turfs (which was held to result in the receipt of income). *Paget's* case was decided on a different basis but the observations of Lord Romer and the basis on which he decided *Paget's* case are consistent with my analysis.

H *Premiums on leases*

32 A premium payable by a tenant on taking a lease has been treated in law as a sum paid in order to obtain a capital asset in the form of a new lease, and thus as constituting a capital payment by the tenant. This is so even if the amount of the premium has been calculated by reference to the

prospective rent. In *Strick v Regent Oil Co Ltd* [1966] AC 295 premiums were payable on leases at nominal rents from 5 to 21 years in duration granted by a petrol retailer in favour of a petrol supplier, who paid the premiums. The supplier then granted subleases of the premises back to the retailer with restrictive covenants, or a "tie", in favour of the supplier. Lord Reid observed, at p 325, that premiums paid for leases had always been regarded as capital. However, he left open the position where a premium had been paid for a very short lease, say two or three years. Apart from Lord Reid, the members of the House gave comparatively little consideration to the possibility that on a special set of facts a premium for a lease might actually constitute income. Lord Morris of Borth-y-Gest, at p 334, considered that a tie for five or even fewer years was a capital asset but Lord Upjohn, at p 346, and Lord Pearce, at p 336, both thought that periods of more than five years were necessary in the case of mere contractual ties. The House of Lords held that the premiums constituted capital payments by the petrol supplier. The essential point, common to both premiums and ties, was that, for the payments to be treated as capital, an asset or advantage of a capital nature would have to be created.

33 In *Comr of Inland Revenue v Wattie* [1999] 1 WLR 873 the question was whether a reverse premium was capital or income in the hands of the lessee who received it. It was held to be the former on the basis that it was a once and for all payment for taking on an onerous lease for a substantial period. I should mention an additional submission which Mr Goldberg made on the basis of this case, namely that, if in all the circumstances a receipt could not be attributed to the profits of a particular year, then in the absence of special statutory provisions, it had to be left out of account: see the speech of Lord Nolan, at p 882, giving the advice of the Privy Council. However, Mr Henderson submitted, in my judgment correctly, that if in this case the proceeds are income, they are properly attributable to the year of receipt.

34 Mr Goldberg has argued that JLP could certainly have ensured that the proceeds were a capital receipt if it had granted a five-year lease to Rabobank at a nominal rent and with a premium equal to the amount of the proceeds. That submission seems to me, with respect, not to take the matter much further since the transaction with which we are concerned is a very different one. To treat the proceeds in the present case as if they were a premium paid on the grant of the lease would in my judgment be to do the very thing that both counsel are agreed the court cannot do, that is recharacterise the transaction as something other than that effected by the assignment.

35 Where a premium is paid, a capital asset is created, viz a lease. Such an event is to be distinguished from the payment of a lump sum representing the accumulation of payments for which a tenant is already liable (even if not immediately) in return for day to day use of demised premises. This is a distinction made by Lord Wilberforce, speaking of the premiums in issue in *Strick v Regent Oil Co Ltd* [1966] AC 295, 349:

> "They were lump sums, paid at the start of the transactions to procure the immediate emergence of an asset or advantage, enjoyment of which was secured for a period. They were not, and did not represent the aggregation of, current payments made for the day-to-day use of or continuation of an advantage."

A 36 For myself, I have found the cases on premiums to have been of limited assistance on this appeal. The treatment of premiums for fiscal purposes is now subject to specific statutory provision: see section 34 of the Income and Corporation Taxes Act 1988 referred to below. Moreover, the payment of a premium must be distinguished from an advance payment of rent, which gives rise to an income receipt in the hands of the lessor: *Greyhound Racing Association (Liverpool) Ltd v Cooper* [1936] 2 All B ER 742. The lessor who receives a premium is able to say that he has realised an element of the capital value of his property by granting a new lease. The difference between that and the present case is that JLP has not, in my judgment, in reality realised an element of the capital value of its property but has disposed of some of the income which it produces or will produce in future. Thus the present case is not on all fours with the cases on C premiums. Moreover, examination of the cases serves to illustrate and emphasise the importance of analysing the source, as a practical business matter, of the payment in issue. In *Strick's* case the source was the transaction of the grant of a lease. In the present case the source is the assignment. This involves the disposal of choses in action but that is a juristic classification, not a classification from a commercial or business point of view, and juristic distinctions carry little weight in this field.

D 37 I would regard as a capital asset an asset or advantage which is held for the enduring benefit of the taxpayer's business, here property investment. I take this meaning of capital asset from the celebrated test of capital expenditure set out by Viscount Cave LC in *British Insulated and Helsby Cables Ltd v Atherton* [1926] AC 205, 213–214:

> "But when an expenditure is made, not only once and for all, *but with* E *a view to bringing into existence an asset or an advantage for the enduring benefit of a trade*, I think that there is very good reason . . . for treating such an expenditure as properly attributable not to revenue but to capital." (Emphasis added.)

Compensation cases

F 38 It will be recalled that in *Glenboig Union Fireclay Co Ltd v Inland Revenue Comrs* 1922 SC(HL) 112 the taxpayer received compensation for inability to continue working its fireclay fields. This was held to be a capital sum. The principle that has been applied is that if the compensation was for loss of income, the receipt is treated as income and likewise if the compensation was for loss of a capital asset (as in the *Glenboig* case), the G receipt is treated as a capital receipt: see per Diplock LJ in *London and Thames Haven Oil Wharves Ltd v Attwooll* [1967] Ch 772, 815. Thus, for example, in *Deeny v Gooda Walker Ltd (No 2)* [1996] 1 WLR 426 damages awarded to Lloyd's names against their managing agents for the negligent conduct of the names' underwriting affairs for several years of account were held to be taxable as income, since the damages replaced the losses which the names in fact made. So, too, where a timber merchant lost his stock in trade H of timber through fire, the insurance moneys were income, per Viscount Dunedin in *J Gliksten & Son Ltd v Green* [1929] AC 381, 385:

> "The whole point is that the business of the company is to buy timber and to sell timber and when they sell timber they turn it into money. This particular timber was turned into money, not because it was sold, but

because it was burned and they had an insurance policy over it. The whole question comes to be whether it is turnover in the ordinary course of their business. I think it was. They had that amount of timber, which they got rid of and for which they got a certain price, and then they could begin again." A

39 In *Raja's Commercial College v Gian Singh & Co Ltd* [1977] AC 312 the Privy Council (Lord Diplock, Lord Simon of Glaisdale and Lord Fraser of Tullybelton) held that the principle summarised above from the *Attwooll* case [1967] Ch 772 applied equally to traders and investors. The tax treatment of compensation did not depend on whether the recipient was a trader rather than an investor, but on the "essential character" of the compensation. "The essential character" of the compensation directs one back to examine the matter for which the compensation is being paid. B

40 The compensation cases are all part of a piece with the other cases that I have cited. The theme in all these cases is that the nature of the asset, from which the payment in issue is derived, has a strong influence on the characterisation of that payment as income or capital. C

Sale of rights to income

41 I have already dealt with *Paget v Inland Revenue Comrs* [1938] 2 KB 25, in which Miss Paget sold interest coupons and was held not to have received income arising from securities. The interest coupons must have been detachable from the bonds and thus constitute the entirety of Miss Paget's property in the interest coupons. In *Inland Revenue Comrs v McGuckian* [1997] 1 WLR 991 the question in issue was in essence whether a shareholder, Shurltrust, by assigning the right to any dividends paid in 1979 by a company called Ballinamore to another company called Mallardchoice for a lump sum had disposed of a right to income and received capital in circumstances where Mallardchoice immediately paid 99% of the dividend it received back to Shurltrust. The case was decided on *Ramsay* principles (see *W T Ramsay Ltd v Inland Revenue Comrs* [1982] AC 300) and the capital/income distinction was not argued. Accordingly the *McGuckian* case is not authority for the proposition that the sale of the right to receive income indicates that the proceeds constitute a capital receipt. Even so, Lord Browne-Wilkinson expressly proceeded on the basis that that was the prima facie position, and, as Mr Goldberg observed, the issue in that case would not have arisen as it did if it had not been conceded that this sale of the right to future dividend produced a capital sum. This, he submitted, supports his argument that sale of the right to income results in a capital receipt. However, it is not clear why the revenue took that view in the *McGuckian* case. The share capital of Ballinamore was £500 and the dividend was nearly £500,000, being the totality of its distributable reserves (no mention being made of any undistributable reserves). Factually, therefore, the dividend in issue in the *McGuckian* case was quite different from the recurring rents which JLP disposed of in this case. In any event, Lord Steyn made it clear in the *McGuckian* case [1997] 1 WLR 991, 1002C–D, that he would not have regarded the money in question as capital in any event and Lord Cooke of Thorndon, at p 1003D, did not regard it as axiomatic that the proceeds of sale of a right to receive a dividend would be capital, although he expressed the view that in many cases this would be so. D E F G H

A 42 In *MacNiven v Westmoreland Investments Ltd* [2003] 1 AC 311, 332 Lord Hoffmann, with whom the other members of the House agreed, set out the approach to the interpretation of tax statutes. He drew an analogy with the distinction between income and capital where it is well established, as shown above, that those terms have a commercial meaning and whether particular receipts constitute income or capital is to be ascertained from a B practical and business point of view. His speech also gives important guidance as to how to apply a term which has a commercial meaning. He said, at pp 328–329, para 39:

> "My Lords, I venture to suggest that some of the difficulty which may have been felt in reconciling the *Ramsay* case [1982] AC 300 with *Inland Revenue Comrs v Duke of Westminster* [1936] AC 1 arises out of an C ambiguity in Lord Tomlin's statement that the courts cannot ignore 'the legal position' and have regard to 'the substance of the matter'. If 'the legal position' is that the tax is imposed by reference to a legally defined concept, such as stamp duty payable on a document which constitutes a conveyance on sale, the court cannot tax a transaction which uses no such document on the ground that it achieves the same economic effect. On the other hand, if the legal position is that tax is imposed by reference to a D commercial concept, then to have regard to the business 'substance' of the matter is not to ignore the legal position but to give effect to it."

Lord Hoffmann continued, at pp 332–333:

> "50. The distinction between commercial and legal concepts has also been drawn in other areas of legislation. So, for example, the term E 'financial assistance' in section 151 of the Companies Act 1985 has been construed as a commercial concept, involving an inquiry into the commercial realities of the transaction: see *Burton v Palmer* [1980] 2 NSWLR 878, 889–890, *Charterhouse Investment Trust Ltd v Tempest Diesels Ltd* [1986] BCLC 1. But the same is not necessarily true of other terms used in the same section, such as 'indemnity'. As Aldous LJ said in *British and Commonwealth Holdings plc v Barclays Bank plc* [1996] F 1 WLR 1, 14: 'It was submitted that as the words "financial assistance" had no technical meaning and their frame of reference was the language of ordinary commerce, the word "indemnity" should be similarly construed. The fallacy in that submission is clear. The words "financial assistance" are not words which have any recognised legal significance whereas the word "indemnity" does. It is used in the section as one of a G number of words having a recognised legal meaning.' I would only add by way of caution that although a word may have a 'recognised legal meaning', the legislative context may show that it is in fact being used to refer to a broader commercial concept.
>
> "51. In the *McGuckian* case [1997] 1 WLR 991 a Republic of Ireland company called Ballinamore had substantial distributable reserves. The shareholders, Mr and Mrs McGuckian, wanted to receive this money but H not to pay income tax on the dividend. So they entered into a scheme by which they first transferred their shares to an offshore trustee called Shurltrust. By a series of preplanned transactions, it then assigned the right to receive the dividend to a United Kingdom company called Mallardchoice in consideration of the payment of a sum equal to 99% of

the expected dividend. Ballinamore then declared the dividend and paid it to Mallardchoice, which immediately paid 99% to Shurltrust.

"52. The statutory question was whether Shurltrust had received income or capital. If it was income, the effect of various tax avoidance provisions concerning the transfer of assets abroad was that the payment would be deemed to be income of the McGuckians. If it was capital, the McGuckians would not be liable for tax. The McGuckians said that if Shurltrust had simply received the dividend, it would of course have been income. But Shurltrust did not receive the dividend. It received a payment from Mallardchoice which was a capital payment for an assignment of its right to income.

"53. The Inland Revenue's argument, relying upon the formulation in *Furniss v Dawson* [1984] AC 474 was that the assignment should be disregarded. The Northern Ireland Court of Appeal said (not, if I may respectfully say so, without justification) that one could not simply 'disregard' the assignment. The payment of the money by Mallardchoice to Shurltrust was the consideration for the assignment and an integral part of that transaction. If the assignment had to be disregarded, one could not explain how Shurltrust had received any money at all.

"54. . . . In the *McGuckian* case [the question] was the nature of the payment received by Shurltrust—capital or income? . . . The question was not whether the assignment should be disregarded but whether, from a commercial point of view, it amounted to an exchange of income for capital. Such exchanges usually have a commercial reality: the purchase or sale of an annuity, for example, is an exchange of capital for an income stream, involving a transfer of risk. But the transaction in the *McGuckian* case was nothing more than an attempt to relabel a sum of money. The fact that the assignment had no commercial purpose did not mean that it had to be disregarded. But it failed to perform the alchemy of transforming the receipt of a dividend from the company into the receipt of a capital sum from someone else. For the purpose of the fiscal concept at stake, namely the character of the receipt as income derived from the company, it made no difference."

Lord Hoffmann continued, at pp 334–335:

"58. The limitations of the *Ramsay* principle therefore arise out of the paramount necessity of giving effect to the statutory language. One cannot elide the first and fundamental step in the process of construction, namely to identify the concept to which the statute refers. I readily accept that many expressions used in tax legislation (and not only in tax legislation) can be construed as referring to commercial concepts and that the courts are today readier to give them such a construction than they were before the *Ramsay* case. But that is not always the case. Taxing statutes often refer to purely legal concepts. They use expressions of which a commercial man, asked what they meant, would say 'You had better ask a lawyer'. For example, stamp duty is payable upon a 'conveyance or transfer on sale': see Schedule 13, paragraph 1(1), to the Finance Act 1999. Although slightly expanded by a definition in paragraph 1(2), the statutory language defines the document subject to duty essentially by reference to external legal concepts such as 'conveyance' and 'sale' . . .

A "59. Even if a statutory expression refers to a business or economic concept, one cannot disregard a transaction which comes within the statutory language, construed in the correct commercial sense, simply on the ground that it was entered into solely for tax reasons. Business concepts have their boundaries no less than legal ones . . .

"60. Likewise the use of business concepts like 'income' and 'capital' may give the taxpayer a choice of structuring a commercial transaction so B as to come within one concept or the other. As Lord Greene MR said in a celebrated passage in *Inland Revenue Comrs v Wesleyan and General Assurance Society* [1946] 2 All ER 749, 751: 'In dealing with income tax questions it frequently happens that there are two methods at least of achieving a particular financial result. If one of those methods is adopted, tax will be payable. If the other method is adopted, tax will not be C payable. It is sufficient to refer to the quite common case where property is sold for a lump sum payable by instalments. If a piece of property is sold for £1,000 and the purchase price is to be paid in ten instalments of £100 each, no tax is payable. If, on the other hand, the property is sold in consideration of an annuity of £100 a year for ten years, tax is payable. The net result, from the financial point of view, is precisely the same in D each case, but one method of achieving it attracts tax and the other method does not.'

"61. It follows that a transaction which, for the avoidance of tax, has been structured to produce, say, capital, and does produce capital in the ordinary commercial sense of that concept (unlike the payment in *Inland Revenue Comrs v McGuckian* [1997] 1 WLR 991) cannot be 'recharacterised' as producing income: see *Comr of Inland Revenue v* E *Wattie* [1999] 1 WLR 873."

43 Lord Hoffmann's use of the term "commercial reality" in *MacNiven's* case is reminiscent of his approach to the expression "financial assistance" in what was then section 54 of the Companies Act 1948. "The words have no technical meaning and their frame of reference is in my judgment the language of ordinary commerce. One must examine the F commercial realities of the transaction": *Charterhouse Investment Trust Ltd v Tempest Diesels Ltd* [1986] BCLC 1, 10, per Hoffmann J—a case cited by Lord Hoffmann in *MacNiven's* case [2003] 1 AC 311, 332, para 50.

44 However, in *MacNiven's* case Lord Hoffmann was only dealing with the capital/income question by analogy and he did not seek to examine the authorities on that question. The test of commercial reality does not G represent a completely new approach to that distinction, but constitutes the most recent and authoritative guidance on the capital/income question. Income and capital are commercial concepts. To give effect to them, the court must have regard to the "business 'substance' of the matter": see per Lord Hoffmann in *MacNiven's* case [2003] 1 AC 311, 328–329, 332, 333, 334, paras 39, 40, 50, 54, 58.

H *Points arising from the English and Scottish cases*

45 Certain points can be derived from all these cases. The first point is that every case depends on careful examination of the particular circumstances. The decision whether a receipt (or item of expenditure) is capital or income is highly fact-sensitive. Second, "capital" and "income"

are commercial concepts. The test is one of commercial reality and technical juristic distinctions are of little relevance. The court must pay regard to the transactions which took place, but it matters little that (in this case) the assignment effected the transfer of an interest in land or the disposal of chose in action, whether at law or in equity. Third, the underlying asset from which the sum is derived may have a large influence on whether the payment is capital or income. The compensation cases also show that the interposition of a third party (normally an insurer or tortfeasor) does not alter the importance of the source of the payment in issue in this sense. Fourth, the court asks whether as a matter of practical and business reality the asset transferred was (so to speak) the fruit or the tree on which the fruit hung. Accordingly, a distinction is drawn between fruit and the fruit-producing element of an asset. Fifth, the sale of part of the asset does not always result in a capital receipt: see, for example, *Withers v Nethersole* [1948] 1 All ER 400 and *Lowe v J W Ashmore Ltd* [1971] Ch 545. Sixth, the court also examines carefully the nature of the payment. But one must be wary of relying solely on the nature of the payment. The mere fact that it is a lump sum does not mean that it is capital. For example, a person may write an article for the local newspaper and transfer to it the copyright to it and receive in exchange a lump sum and yet be held to have received income. On the other hand, a capital asset may be realised by instalments: see *Inland Revenue Comrs v Wesleyan and General Assurance Society* [1946] 2 All ER 749, 751 per Lord Greene MR, cited by Lord Hoffmann in the *MacNiven* case [2003] 1 AC 311, 335, para 60, above. Recurrence is relevant in that it generally signifies income but, as in the example just given about writing an article for a newspaper, income need not be recurrent. The mere fact that a receipt is a sum calculated by reference to trading profits does not mean that it is income, though it will often be. None of the factors identified in this paragraph is decisive, but they may indicate that a particular outcome is to be preferred.

The Australian cases

46 We were shown two Australian decisions, *Comr of Taxation of the Commonwealth of Australia v Myer Emporium Ltd* (1987) 163 CLR 199, a decision of the High Court of Australia (Mason ACJ, Wilson, Brennan, Deane and Dawson JJ), and *Henry Jones (IXL) Ltd v Federal Comr of Taxation* (1991) 102 ALR 1, a decision of the Federal Court of Australia on appeal from Sweeney J. In the former case the issue was whether the sale of future interest payable on a loan for more than seven years in exchange for a lump sum constituted income or capital. The High Court distinguished *Paget's* case and held 163 CLR 199, 218–219:

> "If a lender sells his mere right to interest for a lump sum, the lump sum is received in exchange for, and ordinarily as the present value of, the future interest which he would have received. This is a revenue not a capital item—the taxpayer simply converts future income into present income: see *Comr of Internal Revenue v P G Lake Inc* (1958) 356 US 260, 266–267 . . . In *Paget* [1938] 2 KB 25 . . . the coupons had come to represent, like a contract to pay an annuity, the sole source of the expected payments. Lord Romer, at pp 44–45, drew that analogy, treating the sale for a lump sum of an annuity as an instance of a sale of a

A right to receive income in the future, the proceeds of which are not treated as income. Unlike the sale of the coupons in *Paget*, the sale of a right to interest severed from the debt is not a sale of a tree of which the future payments are the fruit. The present case may thus be distinguished from the view of the facts which was the foundation of the decision in *Paget*. If *Paget* is not to be distinguished in this way, we should be unable to accept
B its authority for the purposes of the Act."

47 I do not think it is necessary to set out the decision in the *Henry Jones* case 102 ALR 1 as, so far as material, it applied the same principle.

48 These cases are persuasive authority only, but they are consistent with the conclusions reached above. I do not agree with Lightman J [2002] 1 WLR 35, 48 that the conclusions of the High Court are "in total conflict"
C with the observations of Lord Romer, given that Lord Romer was dealing with detachable coupons. The High Court of Australia recognised that the coupons were "the sole source of the possibility of making some money abroad; they were the sole source of the expectation" of the payments offered in lieu by the issuers on their default: see the *Myer Emporium* case 163 CLR 199, 219.

49 The analysis of the High Court of Australia is of benefit because it
D has already had to consider the question with which we are concerned. Dame Sian Elias, Chief Justice of New Zealand, speaking extrajudicially said recently that the advantage of the common law is the provision of reasons for judgments. Lord Rodger of Earlsferry, also speaking recently extrajudicially, has made the same point but emphasised that such reasons are given in the particular context of the jurisdiction in question. However,
E subject to that qualification, reasons given by other courts can cut across international boundaries and enable us to derive valuable assistance from other parts of the common law world. That said, the High Court of Australia's conclusion was in reference to the facts of the case before it and could not be taken for the purposes of English law, even on the assumption that English law is the same, to have established an immutable principle.

F *Statutory provisions*

50 Mr Goldberg submitted that other fiscal legislation, namely sections 34, 56, 56A (as inserted by section 34 of and paragraphs 1 and 6 of Schedule 8 to the Finance (No 2) Act 1992) and 730 of the Income and Corporation Taxes Act 1988, proceeds on the assumption that the price paid for what is in effect the sale of the right to an income is capital and not
G income. Mr Goldberg also relied on the form of the amendments made by section 110 of the Finance Act 2000.

51 Section 34 provides that a proportion of a premium payable under a lease, calculated by applying an algebraic formula, will be treated as rent. As this deals with premium I do not accept the premise of Mr Goldberg's submission that this is the sale of a right to income. Section 56 provides that
H the disposal of the right to interest arising under certificates of deposit or similar transactions will be treated, *if it would not otherwise be treated*, as annual profits. Section 56A extends section 56 to the disposal of the right to receive an amount pursuant to a deposit under a transaction where there is no certificate of deposit but only the right to call for such a certificate. The words italicised make it clear that Parliament did not proceed on the basis

that such sales would necessarily be capital. Section 730 provides that if there is a disposal of the right to interest payable in respect of a security without disposal of the security itself, that interest is deemed to be income of the owner of the security. This section seems to me to be dealing with the attribution of income rather than the income/capital question. A

52 The effect of sections 43A, 43B and 43C of the Income and Corporation Taxes Act 1988, as inserted by section 110 of the Finance Act 2000, is that if the deed of assignment were now executed the proceeds would be treated as income. The new provisions apply, however, only to short-term transactions and Mr Goldberg submitted that the implication is that a disposal of the right to receive income would otherwise be capital. Lightman J agreed with this inference and took "some comfort" from these provisions but concluded that they could not shed much light on the law in force five years previously: [2002] 1 WLR 35, 51–52, paras 41, 42. In my judgment, it would be unsound on this appeal to draw any inference about the application of the capital/income question to transactions outside sections 43A to 43C. Section 43A is structured in such a way that it makes critical the question whether under normal accounting practice the assignor of rents would have to account for the proceeds received under the assignment as a financial liability (in this case, as a loan). Little argument has been addressed in this court on the relevance of the revenue's Financial Reporting Standard 5 on the capital/income question in the absence of section 110. Given that the court is required to examine the transaction from a practical and business point of view, the accounting treatment is one of the relevant circumstances and may provide some assistance to the court: *Strick v Regent Oil Co Ltd* [1966] AC 295, 355, per Lord Wilberforce. Moreover, section 43A only applies to assignments of rent for terms not exceeding 15 years and there could have been separate policy reasons for limiting the new provisions in this way. However that may be, the fact that the legislature proceeded on a particular basis as to the general law cannot in any event preclude the court from declaring what the general law was: *West Midland Baptist (Trust) Association (Inc) v Birmingham Corpn* [1970] AC 874, 898, per Lord Reid. B C D E F

The judge's analysis

53 In his conclusion [2002] 1 WLR 35, 50, para 37, the judge identified five reasons for holding that the receipt of the proceeds of the deed of assignment was capital not income: see paragraph 9 above. His first reason was that of taxpayer choice. However, there is no question as to that in this case and there is no suggestion that the transaction should be taxed as income simply because it could have been effected in a manner that clearly produced income. I agree, but as Lord Hoffmann said in *MacNiven's* case [2003] 1 AC 311, 328–329, para 39, where a commercial concept is used, to have regard to the "business" substance is not to ignore the legal position but to give effect to it. The judge's second reason was that the transaction produced a lump sum (he called it a "capital" sum but that of course was a matter of conclusion) in place of an income stream. However, this is only an indicator and cannot be conclusive in itself: this follows from my analysis of Lord Romer's observations in *Paget v Inland Revenue Comrs* [1938] 2 KB 25 and other cases, above, for example *Nethersole's* case [1948] 1 All ER 400 and the *Greyhound Racing Association* case [1936] 2 All ER 742. G H

A The judge's third reason was that this was not a transaction under which rents were "relabelled". The proceeds under the deed of assignment were a distinct sum paid out of the resources of Rabobank under a transaction which had commercial reality. I accept that Rabobank paid the proceeds out of its own resources. However, the whole transaction must be reviewed to see whether as a matter of commercial reality it produced income rather than capital. In *MacNiven's* case [2003] 1 AC 311 the abortive attempt in the
B *McGuckian* case [1997] 1 WLR 991 to turn income into capital was said to have been nothing more than an attempt at relabelling. The judge concluded (as his fourth reason) that in those circumstances it was not open to the court to recharacterise the proceeds as income. It is not suggested by the revenue that the proceeds should be recharacterised. Recharacterisation would involve disregarding the assignment. Nor does economic equivalence play
C any part in this case: that would involve applying some broad functional test to establish the nature of the proceeds. As his fifth reason, the judge held that there could be an exchange of income for capital even though there was no substantial risk of loss through non-payment and in any event in this case there was the theoretical risk of loss if the John Lewis Group went into insolvent liquidation. The question whether there has been a real transfer of risk is relevant because the court is examining the transaction from the
D business point of view. Transfers of theoretical risk, however, are likely to be of less impact in that context.

Conclusions

54 In my judgment, there are a number of other features of the transaction which must be addressed apart from those enumerated by the
E judge in his conclusion. First, JLP's profits were higher as a result of the assignment than they would have been if the rents had been received in the usual way. This factor provides a real commercial motive for the transaction. However, I do not consider that this alone can resolve the income/capital question. Commercial motive and commercial reality are different concepts. The one can exist without the other. Second, JLP made
F the proceeds available for application by the group in expenditure on capital improvements but the evidence does not in my judgment show that the sale proceeds performed any different function in its economy from that which the rents would have performed if they had been discounted by Rabobank. JLP continued to show the rents as rents receivable in its published accounts. Third, so far as the banking transaction with Rabobank was concerned, it is not possible just to look at the transaction as between Rabobank and JLP.
G The transaction was highly unusual because the tenant was a party to the transaction and gave a guarantee, and as a result the only risk transferred was a theoretical one. If there was default in the payment of any of the assigned rents, Rabobank could recover the relevant part of the proceeds either from JLP or from JL under the guarantee and indemnity. Fourth, the proceeds were substantial, but I do not myself place weight on that in the
H light of the fact that it was negotiated at arm's length.

55 A further, and to my mind significant, point is that the assignment of six rents payable for just over five years was a very short period as compared with the total period of JLP's ownership. JLP was the owner of the freehold or long leaseholds, and, seen in the context of its entire period of ownership, this particular assignment was but a drop in the ocean. The rights of

Rabobank were extinguished at the end of just over five years. Thus JLP did not dispose of anything of an enduring nature and JLP received the properties back entire and intact. The profit-generating quality of the properties was unaffected. In my judgment, the authorities show that those factors take the case outside the realisation principle (as I have termed it) and the analogy sought to be derived from the cases on premiums. As I explain below, the position would probably have been different if the disposal had been of rents for the entire period of its ownership but in any event the revenue do not contend that anything other than a short-term disposal of the rents constitutes income: see paragraph 12 above. Moreover, I would not accept that the fact that the rent paid by JL to JLP would be income in the hands of JLP is irrelevant. In my judgment, the authorities on compensation examined above show that the nature of the asset which gives rise to the payment is a very important factor on the capital/income question: it can give it its "essential character".

56 Each case in this field must depend on its particular circumstances. The duty of the court is to look at all the relevant considerations and weigh them appropriately, applying common sense rather than any fixed legal precept. It is helpful to pose the question which Lord Hoffmann said in *MacNiven's* case ought to have been asked in the *McGuckian* case [1997] 1 WLR 991:

> "The question was not whether the assignment should be disregarded but whether, from a commercial point of view, it amounted to an exchange of income for capital. Such exchanges usually have a commercial reality": [2003] 1 AC 311, 333, para 54.

57 The next question is: what is it that according to Lord Hoffmann will usually have commercial reality? It is not open to the court to disregard the assignment itself Accordingly, in the context of the observation cited in the preceding paragraph, it seems to me that Lord Hoffmann must be referring to commercial reality as an exchange of income for capital or vice versa. For that reason Lord Hoffmann takes the example of the purchase or sale of an annuity (meaning, as I understand it, the exchange of a lump sum for the instalments of an annuity) since instalments of an annuity without doubt provide an income stream in the absence of special circumstances. In this case, therefore, the question is whether the conversion by JLP of the rents into the proceeds has commercial reality as an exchange of income for capital.

58 Unfortunately for the court, the special commissioner [2000] STC (SCD) 494 did not consider this question as he considered that he was bound by *Paget's* case [1938] 2 KB 25 to hold that the proceeds were capital. It would have been helpful to me to have had his assessment of the commercial reality. Mr Haberman's evidence does not assist since the description of the transaction as a financing transaction is equivocal: a transaction could provide financing whether it gave rise to a receipt which was income or capital in nature. I must accordingly do the best I can without assistance. Balancing the relevant factors in this case, the telling factors to my mind are first the nature of the asset transferred and second the nature of the payment.

59 The asset transferred consisted of the six rents plus the rights to recover the same. The rents and those rights were income in the hands of

A JLP. If asked to identify the "business" substance or commercial reality of the matter, a businessman or woman would, as it seems to me, looking at all the circumstances of this transaction, say "JLP has entered into an arrangement with its bankers to discount six rents" or "JLP has entered into arrangements to receive accelerated payments of rent, discounted for early payment" I do not consider that he or she would say: "JLP has sold a part of five of its properties" or "JLP has sold six choses in action under leases B relating to five of its properties". It is true that JLP received a lump sum but that is because it chose to aggregate a number of rentals together and to receive a single accelerated payment for them. In those circumstances the fact that JLP received a lump sum is of less weight. Furthermore, as the authorities make clear, the fact that the assignment transferred an interest in property and choses in action is not a determining consideration since it is C not the juristic classification with which this question is concerned. Moreover, I do not consider that JLP can in any business sense be said to have exploited the ability of the properties to generate rents after the period covered by the assignment.

60 Mr Goldberg submitted that by assigning the rents JLP has parted with part of its property and therefore, it follows, JL received a capital D payment. I do not accept this argument. The assignment of the right to recover the rent did not impede JLP from recovering rents in the future. Like the turfs in *Lowe v J W Ashmore Ltd* [1971] Ch 545, the rents and the choses in action are property that will recur. To focus on the once and for all disposal of the particular six rents is to disregard the full picture. The assignment does not result in a disposal of any income-producing part of the property as opposed to income, and this is required by the realisation E principle as I have termed it. The resultant diminution in the market value of the properties was only a temporary fluctuation in value. A reduction in capital value does not in my judgment denote that JLP made a disposal of a capital asset. As investment properties, the properties will have been valued on a yield basis, and this is bound to be temporarily affected by the assignment. However, that factor does not make the assigned rents part of a F capital asset. They remain income. The court can take into account also the fact that the fluctuation in value is only temporary. When looking at the business substance of the matter the court is not restricted to looking at the dip in value at the date of the assignment and not what happened thereafter. If the position were otherwise, the fact that Mallardchoice paid 99% of the dividend back to Shurltrust in the *McGuckian* case [1997] 1 WLR 991 might have to have been disregarded. In short, save on a short-term basis, the G income-earning capacity of the properties with respect to subsequent rents was unaffected.

61 As to the nature of the payment, it is clear from the authorities that the fact that it was calculated by reference to rents does not make it income. Here, however, rentals were not simply a measuring stick: they were the asset in reality realised by the whole transaction. The amount of the H proceeds represents the present value of the rent. Moreover, in a practical sense, risk of default was not transferred. If the tenant defaulted, Rabobank had the right to put JLP back in the position it would have been in if the rent on which default had occurred had not been assigned. Because risk was not transferred, JLP continued to account for the rents in its statutory accounts as and when received rather than taking credit in its accounts for the

proceeds in the year of receipt. Significantly it included the receipts as rents and not by any other description. A

62 What then is the business substance or commercial reality of the "proceeds"? In my judgment, in all the circumstances of this case, they must take their capital/income colour from the rentals which they represented. Instead of receiving rents as and when they fell due, JLP simply realised them in a different way. The assignment is not to be disregarded. The fact, however, that the process involved the necessary interposition of a bona fide third party bank, which received the rents and paid the proceeds out of its own resources should not deflect the eye from the real nature of the transaction. A third party was interposed in the *McGuckian* case and in the compensation cases. The commercial reality of the assignment is that JLP discounted the rents. It seems to me to follow that, whatever terminology the parties used in the deed of assignment, the commercial reality of what JL received was discounted rents and hence income. The transaction effected no change in JLP's programme of investment in properties. JLP merely realised the six rentals in question in another way. What the transaction was intended to effect, from a practical and business point of view, was an accelerated payment of the rents, discounted for early payment. I cannot see that the proceeds should be capital any more than they would have been if JLP has sold the rents each year for six years by six separate transactions. B C D

63 I would accept that the position would probably have been different if JLP had itself been entitled to a headlease for 50 years and had sold the rents for the entire period of its headlease in one transaction. If that is right, then depending on the circumstances an assignment of rents occurring over a lesser period than the period of ownership may suffice. However, all that happened here was, comparatively speaking, but a ripple in the rental stream, and, when the full circumstances are examined, the assignment did not as a commercial matter change the nature of the receipt from income to capital. E

64 For all these reasons I consider that the proceeds of the assignment were, as a matter of business and commercial reality, income and not capital. In the circumstances, I would allow this appeal. F

The subsidiary question—the head of charge

65 Schedule A brings into charge annual profits "in respect of any such rents or receipts as follows . . . receipts arising to a person from or by virtue of his ownership of an estate or interest" in land. Mr Goldberg submits that the assignment was the sale of an interest in land. I agree with the judge that these words, in particular the words "by virtue of", are wide and general and cover the proceeds on the basis they are income. G

66 Case VI of Schedule D charges to tax annual profits or gains not falling under any other case of Schedule D and not charged by virtue of Schedule A, C or E. Mr Goldberg argued that such profits and gains must be profits or gains of the same genus as profits or gains arising under other heads of charge, relying on a passage from the judgment of Blackburn J in *Attorney General v Black* (1871) LR 6 Exch 308, 309, quoted by Viscount Dunedin in *Jones v Leeming* [1930] AC 415, 422, that to be chargeable under general sweeping provisions in Schedule D (as it then stood) the receipt had to be of the same nature and kind as had been previously H

A mentioned. This was applied by Viscount Dunedin in the context of a question as to income or capital. In my judgment, the proceeds would fall within the genus established by Schedule A since it arises from a sale of rents arising from an estate or interest in land. Accordingly, on the basis that the proceeds are income, and if I am wrong in concluding that the proceeds are taxable under Schedule A, in my judgment they are brought into charge by Case VI of Schedule D.

B 67 Accordingly, I would dismiss the respondent's notice.

DYSON LJ

Introduction

68 The question whether a payment is to be regarded as capital or

C income has troubled the courts for a very long time. There are statements of the highest authority which indicate that classification cannot be made by the application of something akin to a simple litmus test. Various guidelines have been given from time to time. But it has been repeatedly emphasised that much depends on the nature of the transaction and the matrix in which it is set. Thus, in *Van den Berghs Ltd v Clark* [1935] AC 431, 438–439 Lord

D Macmillan said:

> "While each case is found to turn upon its own facts, and no infallible criterion emerges, nevertheless the decisions are useful as illustrations and as affording indications of the kind of considerations which may relevantly be borne in mind in approaching the problem."

69 In *Strick v Regent Oil Co Ltd* [1966] AC 295, 313 Lord Reid said:

E

> "So it is not surprising that no one test or principle or rule of thumb is paramount. The question is ultimately a question of law for the court, but it is a question which must be answered in light of all the circumstances which it is reasonable to take into account, and the weight which must be given to a particular circumstance in a particular case must depend rather on common sense than on strict application of any single

F legal principle."

70 Similar observations were made by Lord Morris of Borth-y-Gest, at p 328B, and Lord Upjohn, at pp 343E and 345C. Common sense is a necessary tool for any judge to use in reaching a decision. But it is not a sufficient one. The authorities do provide some assistance in pointing the way to finding what circumstances are relevant. As Lord Wilberforce

G pointed out in *Strick's* case, in the course of the numerous decisions which have distinguished between capital and revenue expenditure in relation to widely different trades and varying circumstances, certain "tests" have emerged. But as he said, at p 348:

> "These may be useful, so long as it is recognised that they have emerged a posteriori from the facts of a given situation and that they may

H not always be suitable as guiding lines in other situations. I begin by asking two questions, which may be said to be generally relevant: what is the nature of the payment, and for what was the payment made? These, together with a third question, namely, how that, for which the payment was made, was to be used, were stated by Dixon J in his classic judgment

in *Sun Newspapers Ltd v Federal Comr of Taxation* (1938) 61 CLR 337, 363." A

71 Dixon J said, at p 363, there were

> "three matters to be considered, (a) the character of the advantage sought, and in this its lasting qualities may play a part, (b) the manner in which it is to be used, relied upon or enjoyed, and in this and under the former head recurrence may play its part, and (c) the means adopted to obtain it; that is, by providing a periodical reward or outlay to cover its use or enjoyment for periods commensurate with the payment or by making a final provision or payment so as to secure future use or enjoyment." B

72 The discussions by Lord Wilberforce in *Strick's* case [1966] AC 295 and Dixon J in the *Sun Newspapers* case 61 CLR 337 were in the context of cases about expenditure. The present case concerns the classification of a receipt. I accept that a payment which is properly classified as capital expenditure is not necessarily classifiable as a capital receipt in the hands of the payee. Nevertheless, it was common ground before us that the parallels are sufficiently close for the guidance afforded by the expenditure cases to be of assistance in determining whether a receipt is capital or income. C D

73 One final introductory observation. In relation to the classification of expenditure, Dixon J said in *Hallstroms Pty Ltd v Federal Comr of Taxation* 72 CLR 634, 648:

> "What is an outgoing of capital and what is an outgoing on account of revenue depends on what the expenditure is calculated to effect from a practical and business point of view, rather than upon the juristic classification of the legal rights, if any, secured, employed or exhausted in the process." E

It is common ground that this approach should be applied equally to the classification of receipts.

Strick's case in more detail F

74 It is worth examining *Strick's* case [1966] AC 295 in a little detail, because in my view the speeches contain material which is of some relevance to the question that arises on this appeal. In *Strick* what was at issue was whether premiums paid for four leases for terms of years at a nominal rent were payments of a capital or revenue character. The leases were part of a series of transactions whereby "ties" were created between an oil company and service station retailers. The leases ranged in length from five to 21 years. Lord Reid said, at p 325, that he had no doubt that the lump sums paid for the 21-year ties could not be treated as revenue outgoings even if there were no lease or sublease. The very length of the period of the arrangement for which the lump sums were paid indicated that they were capital payments. With regard to the ten-year and five-year leases, he said that the fact that there had been the grant of leases and subleases was highly relevant and that this pointed to the lump sums as being capital payments. He continued, at pp 325–326: G H

> "Premiums paid for leases have always been regarded as capital but we were not referred to any case where a premium had been paid for a very

A short lease—say two or three years, and I do not wish to decide whether even in such a case a premium would necessarily be treated as a capital outlay. But I am satisfied that the weight of this factor in the present cases is sufficient to turn the scale if otherwise there were doubt, and I would therefore hold that in each of the four cases the lump sums paid by the appellants cannot be allowed as revenue outgoings."

B 75 Lord Morris of Borth-y-Gest agreed that each of the lump sums was a capital payment, at p 334:

"Aided by the word pictures or descriptions of a capital asset which the decided cases contain I consider that a tie of the kind now being examined is a capital asset. If a lump sum is paid for such a tie for five years (or for a lesser number of years) it would give a false and unreal picture if the C whole sum were debited to the profit and loss account for the first year or for the year in which the payment was made. If it is said to be hard that no part of the lump sum can be a debit in the profit and loss account that is merely to voice a regret that there is no statutory provision which enables periodic allowances to be made. That however is not a matter for the courts."

D 76 Lord Pearce said, at p 336E, that the fact that an interest in land was being acquired pointed strongly to a capital expenditure. But he also said that, if the premiums had been paid without any acquisition of an interest in land, they would have been of a revenue nature where the transaction was only for a five-year period; but would probably have been capital where the transactions were for a 21-year period, because such transactions would E have acquired "a more enduring and structural quality".

77 Lord Upjohn said, at p 341B–G, that it was axiomatic that a premium for the acquisition of a lease was a capital payment regardless of the duration of the lease. He then went on to consider what the position would have been if there had been no acquisitions of leases, but merely ordinary trading contracts for the same considerations and for the same terms of years as were the subject of the leases. He emphasised that the case F concerned trading contracts and that for that reason the payments were not lightly to be held to be capital. But, at pp 345–346:

"the amount of the payment and the length of the tie are important elements among all the other relevant facts. I part company at once with the submissions of counsel on both sides on the one hand that a lump sum payment for a tie for more than an annual accounting period is necessarily G capital and, on the other, that it is a trading expense and the length of the tie is utterly immaterial save as a factor in calculating the anticipated gallonage and so the amount of the lump sum payment. The lump sum payments here are large. But one must not attribute to that too much importance because after all the lump sum payment is calculated on the basis that it represents no more than one penny per gallon on the expected H sales over the length of the tie. So I approach this matter as one of judicial common sense . . ."

78 He concluded his speech, at p 346, by looking at each of the four transactions without regard to the fact that they had included the acquisitions of leases. Of the 21-year ties he said that "to pay substantial

sums for a tie for as long as 21 years is quite plainly, as a matter of common sense, a tie which must be described as of a capital nature", so that the sums paid for these were to be regarded as capital. The length of the five-year tie put it into the "character of a merely long term trading contract" which would have been an ordinary trading expense. The ten-year tie was in the judgment of Lord Upjohn a "borderline case".

79 Finally, Lord Wilberforce. He was not willing to decide the appeal on the narrow ground that, in relation to an asset so concrete as a lease, at any rate when the term of the lease amounts to five years or more, the test of "durability" is satisfied: p 351G. He considered that the test should not be affected by whether a lease had or had not been granted as part of the consideration for the payment. In his view, there are cases where the transience of the asset acquired is relevant in determining whether it is capital in nature: it depends on the nature of the asset. In the instant case he thought, at p 354:

> "Here the nature of the payments—lump sums—the nature of the advantages obtained—security in respect of the placing of orders for a period—the substantial periods involved, the shortest being a period of five years, more than adequately establish the expenditure as made for the acquisition of capital assets."

He went on to say that he could see no logical basis for saying that 21 or 10 years was good enough to qualify as capital, but three or five years was too short, or for saying that five years or three years may be long enough when there is a lease, and not long enough where there is merely a personal covenant.

Indicia of a capital payment

80 I would identify the following factors in a case such as the present as being relevant to the question whether a payment is capital or income. I emphasise "such as the present" because the guidance derived from cases dealing with one situation may have little application to a wholly different situation. The first factor is duration. If what is disposed of is long-lasting, it is more likely to be a capital asset than if it is something which is evanescent. The cases show that an asset which has an enduring or long-lasting quality is likely to be regarded as a capital asset, and payment received for its acquisition a capital receipt. The converse may not, however, be true. As Lord Wilberforce pointed out in *Strick's* case [1966] AC 295, 353F, if on a consideration of the nature of the asset in the context of the trade in question it is seen to be appropriate to classify it as fixed rather than as circulating capital, "the brevity of its life is an irrelevant circumstance". Context is, therefore, all important. But in the context of the disposal of the right to receive income for a lump sum, the period over which the income is receivable is relevant to the proper classification of the payment for tax purposes. The majority of their Lordships in *Strick's* case considered that, if a premium had been paid for a tie in respect of a period of 21 years, but without the grant of a lease, the payment would have been capital.

81 Secondly, the value of the asset assigned is also a relevant factor: see, for example, per Lord Upjohn in *Strick's* case, at pp 345–346.

82 Thirdly, the fact that the payment causes a diminution in the value of the assignor's interest is material. In my view, there is nothing in the

A authorities to indicate that, unless there has been a permanent impairment of the value of the property, the payment cannot be capital. It is true that in some of the cases, the fact that the value of the property was permanently diminished was regarded as pointing decisively towards the conclusion that the payment was capital. Examples of these are the cases grouped together by Arden LJ as illustrative of what she terms "the realisation principle": B *Glenboig Union Fireclay Co Ltd v Inland Revenue Comrs* 1922 SC(HL) 112, *Earl Haig's Trustees v Inland Revenue Comrs* 1939 SC 676, *Withers v Nethersole* [1948] 1 All ER 400 and *McClure v Petre* [1988] 1 WLR 1386. But these authorities do not support the corollary that, absent a permanent diminution in the value of the property, the payment cannot be capital. Indeed, in *Nethersole's* case in the Court of Appeal Lord Greene MR made the very point [1946] 1 All ER 711, 715:

C

"One might perhaps have expected that where a piece of property, be it copyright or anything else, is turned to account in a way which leaves in the owner what we may call the reversion in the property so that upon the expiration of the rights conferred, whether they are to endure for a short or a long period, the property comes back to the owner intact, the sum paid as consideration for the grant of the rights, whether consisting of a D lump sum or of periodical or royalty payments, should be regarded as of a revenue nature."

83 He continued, at p 715:

"A principle on some such lines as these would not, we think, be out of accord with the popular idea of the distinction between capital and E income. But it is not, we think, open to this court to adopt it as in itself affording a sufficient test . . ."

84 The fact that the diminution is not permanent is not fatal to the classification of the payment as capital. Clearly, if the diminution is permanent, that will suggest strongly that the payment is capital. But the converse is not true. Otherwise, it is difficult to see how the receipt of a F premium for the grant of a lease at a nominal rent can ever be capital, because the grant of such a lease diminishes the value of the landlord's reversion, but only for the duration of the lease. And yet it is plain that the premium for the grant of such a lease for even a short period is capital: *Strick's* case [1966] AC 295. I should add that, in my view, the question whether there has been a diminution of the value of the assignor's interest G should be judged at the date of the assignment. At that date, there has been a diminution in its value: if the assignor were to sell his reversion at that date, he would receive less than if he had not disposed of the right to receive the income from the asset for a period of time.

85 I would, therefore, hold that the fact that the disposal of the asset has caused the value of the assignor's interest to be diminished is a relevant H factor. It seems to me that the amount by which the value of the reversionary interest is diminished is also of some materiality. This should reflect the duration of the asset that has been assigned (my first relevant factor) and its value (my second relevant factor), so that it is likely that the longer the period and the greater the value of the asset, the greater will be the diminution in the value of the reversionary interest. The greater the

diminution in the value of the reversionary interest, the more likely it is that the payment should properly be classified as capital. A

86 The fourth relevant factor is whether the payment is of a single lump sum. If a payment is one of a series of recurring payments made at frequent intervals, it is likely to be income in the hands of the payee. On the other hand, a single lump sum for the once and for all disposal of a particular asset is more likely to be a capital payment. In *Strick's* case [1966] AC 295, 316G, Lord Reid pointed out that ever since *Vallambrosa Rubber Co Ltd v Farmer* 1910 SC 519, "recurrence as against a payment once and for all has been accepted as one of the criteria in a question of capital or income". See also per Lord Morris of Borth-y-Gest [1966] AC 295, 333–334. In some contexts, the fact that payments are recurrent does not tell one anything about whether they are of a capital or revenue nature. For example, in the context of trade, some capital assets only last a very short time and have to be replaced regularly. But a transaction such as a rent factoring agreement with which this appeal is concerned does not involve assets of that kind, any more than did the transactions examined in *Strick's* case. Mr Goldberg submitted that the possibility of recurrence of payments does not, of itself, make a receipt income: there must be more than the mere possibility of recurrence, and in the context of the realisation of an asset, there must also be the existence of a trade relating to the asset: the possibility of recurrence is not relevant to a disposal of a capital asset by a non-trader. I agree that the possibility of recurrence of payments does not, of itself, make a receipt income. It is no more than a relevant factor. But I do not agree that the possibility of recurrence is irrelevant except in relation to the disposal of an asset in the context of trade. It seems to me that the observations of Lord Reid in *Strick's* case about the relevance of recurring payments were not intended to be restricted to the context of trade, but were quite general in their application. B C D E

87 Fifthly, if the disposal of the asset is accompanied by a transfer of risk in relation to it, that tends to suggest that the sum paid for the asset is capital: see per Lord Hoffmann in *MacNiven v Westmoreland Investments Ltd* [2003] 1 AC 311, 333, para 54.

88 I should add that I agree with Arden LJ that this court is not bound by what Lord Romer said in *Paget v Inland Revenue Comrs* [1938] 2 KB 25. I do not propose to deal with the Australian authorities. Suffice it to say that Mr Henderson does not place much reliance on them. F

89 Before I consider whether the judge was right to classify the lump sum payment made in the present case as a capital receipt, I wish to make two further general points. First, the difference between the tree and its fruit. Sir Nicolas Browne-Wilkinson referred to this as "the hackneyed simile" in *McClure v Petre* [1988] 1 WLR 1386, 1390, adding that "it is the distinction between the fruit of the tree, which is income, and the tree itself, which is capital". This imagery undoubtedly serves a purpose in getting across in graphic form the basic idea of the distinction, but it may be more elucidatory in some contexts than in others. Sometimes, it merely begs the question: what is tree and what is fruit? I doubt whether the analogy is helpful in the present case. There is no doubt that if JLP had granted the bank, say, a 15-year lease of the properties for a premium, the premium would have been a capital payment: *Strick's* case [1966] AC 295. But in what sense would that have been a disposal of the tree? And why would it be any more of a disposal G H

A of the tree than the mere assignment of the right to receive rents for 15 years? In both cases, JLP would retain its freehold interest intact, and in both cases the value of its reversion would be (temporarily) reduced by reason of the transaction.

90 The second point is that the way in which the lump sum has been calculated does not shed light on how it should be classified. Thus, the fact that it is equivalent to the value of six years' rent discounted for early receipt B tells one little or nothing about whether the lump sum is capital or income. This point was made in *Strick's* case by Lord Reid, at p 324C, and Lord Wilberforce, at p 349D, "it confuses the measure of the payment with the payment itself", and by Lord Buckmaster in *Glenboig Union Fireclay Co Ltd v Inland Revenue Comrs* 1922 SC (HL) 112, 114–115.

C *The present case*

91 Mr Henderson submits that the payment to JLP was for "the inherently recurrent produce (the rents) of income-producing assets (the properties)": I quote from his skeleton argument. The payment was for future income. It operated as a "substitute" for future rents, and converted future income into present income. What was assigned to the bank was not D the asset comprising the rent-producing properties, but the right to the rent itself. To use the well-known metaphor, the payment to JLP was not for the whole or part of a fruit-bearing tree, but for part of the fruit itself.

92 As I understand it, on this approach, the price paid for an assignment of the right to receive rent can never be capital in the hands of the recipient, no matter how large the sum and how many years' rent is assigned. This conclusion flows inevitably from the fact that the subject matter of the E assignment is the produce of income-producing assets. But in my view this cannot be right. One only has to contemplate a case where the owner of a 30-year lease of property grants a sublease to a tenant for 25 years, and then for a lump sum assigns to a third party the right to receive the whole of the rents payable throughout the 25-year term. Whether one applies an intuitive practical common sense approach, or one applies the five factors that I have F identified, the answer is the same: the lump sum should be regarded as capital. In such a case, the reality is that the owner disposes of a valuable asset for a capital sum, and it is immaterial that, if he had not disposed of the asset but had received the rent every year for 25 years from the tenant, the rent would have been income, and not capital, in his hands. As I have already pointed out, a majority of their Lordships in *Strick's* case [1966] AC 295 thought that a lump sum for tie of 21 years (without the grant of a G lease) would be a capital payment.

93 Mr Goldberg submitted that the payment in the present case was for a part disposal of the properties, and effected a realisation of part of the asset which comprised the properties. Accordingly, the sum received for the realisation was capital. His primary submission is that the sale of the right to income, being the sale of an asset not in the course of trade, must produce H capital, and does not produce income. This submission is of general application and has the merit of simplicity, but I cannot accept it. It does not do justice to the complexity of the issue of classification that is revealed by a consideration of the authorities.

94 Although I cannot accept the full breadth of Mr Goldberg's primary submission, I am of the view that this appeal should be dismissed. The

starting point is that, as between JLP and the bank, the single lump sum payment was not a payment of rent. In this respect, the case is fundamentally different from the prepayment by a tenant of a lump sum representing the discounted value of future rents payable by the tenant under a lease. It is common ground that a lump sum prepayment of rent by a tenant is income in the hands of the landlord. It retains its character as income, notwithstanding that it has been converted into a lump sum. In such a case, there is no disposal by the landlord of an asset. But in the present case, there *was* a disposal of an asset. JLP's right to receive six years' rent was a chose in action which could be assigned for value. In my view, it is irrelevant that *the rent paid by John Lewis to JLP* would be income in the hands of JLP. The relevant question is: what was *the lump sum paid by the bank* in the hands of JLP? It is true that the payment was for future income. But it does not follow that the payment was of the same character as the future income for which it was made. It seems to me that so to characterise the lump sum payment is to make the mistake of confusing the measure by which the payment is calculated with the payment itself: see paragraph 90 above. To apply the words of Lord Wilberforce in *Strick's* case [1966] AC 295, 349B, the lump sum was not, and did not represent the aggregation of, current payments made for the day to say use of or continuation of an advantage. It appeared at first sight to bear the character of a capital payment for an asset.

95 Arden LJ is of the opinion that the source or the nature of the asset from which the payment is derived has a strong influence on the characterisation of that payment as income or capital. Accordingly, in the present case, since the source of the payment is an assignment of income, that points strongly to the payment itself being income too. Reliance is placed on the compensation cases to illustrate this principle. The compensation cases show that where A receives a payment from B to compensate him for the loss of income, then the payment is treated as income. But in my view, the analogy between the compensation cases and the present case is by no means exact. Where A receives compensation for loss of income, the payment is a true substitute for, and therefore equivalent to, income. It fills a hole in his income.

96 Perhaps more importantly, if the source or nature of the asset from which the payment is derived has such a decisive role to play, it is difficult to see why it matters whether what is assigned is five years' rent, 21 years' rent or indeed rent payable over an even longer period. But as I have already stated, it is plain that the length of the period over which the rent is receivable is highly relevant in determining the true character of the payment for tax purposes.

97 I understand Mr Henderson to concede that, if JLP had granted the bank a six-year lease at a nominal rent for a premium, the premium would have been a capital payment. That concession seems to me to be right on the authorities. Looking at the transactions from the bank's point of view, there probably are commercial differences between entering into a new lease for six years at a nominal rent and taking an assignment of the right to receive six years' rent under an existing lease. For example, the new lease may contain repairing and other obligations, although it need not do so. But what matters for present purposes is not how the two transactions are viewed by the bank, but how they are viewed by JLP. From JLP's

A commercial point of view, there are very real similarities between them. In each case, (a) the period is the same; (b) rent which is (or which, in the case of the new lease, absent the premium, would be) payable by the lessee is exchanged for a lump sum which is paid in advance; (c) the reversion is retained by JLP, and (d) the value of the reversion is diminished for (I assume) roughly the same period and, possibly, by roughly the same amount. In each B case, JLP receives a lump sum which represents the value of the rent payable for the properties. The premium for the lease at a nominal rent reflects the amount of the rent that would be payable if the lease were not being granted at a nominal rent, discounted for accelerated payment. So too is the price payable for the assignment of the right to receive the rent actually payable to JLP equal to the value of that rent, discounted for accelerated payment. It is difficult to see in what sense there is any real difference from JLP's C commercial point of view between the two transactions. It is true that the grant to the bank of a six-year lease at a nominal rent for a premium is juristically different from the assignment to it of the right to receive the rents for six years, but as has been made clear in the cases, juristic labels have no part to play in the classification of payments for tax purposes.

98 It follows that I do not share the view of Arden LJ that cases on D premiums have little relevance to the present case. That is not tantamount to treating the proceeds of the assignment as if they are a premium paid on the grant of a lease at a nominal rent, or to recharacterising the transaction as something other than it is. I recognise that there are differences between the two transactions. For example, where a premium is paid, there is created a capital asset, namely a lease, and in some contexts this may be a crucial point of distinction. But, whatever differences there may be between the two E transactions and the two payments in other contexts, I do not consider that, from a practical common sense point of view, the differences are of sufficient substance to justify treating the two payments differently for purposes of taxation. Nor do I consider that a lessor who receives a premium for a five-year lease at a nominal rent is any more able to say that he has realised an element of the capital value of his property than a lessor who has assigned F the right to receive five years' rents. From a practical common sense point of view, in each case the lessor has exchanged the right to receive a stream of payments of rent for a single lump sum, and in each case the transaction will have temporarily reduced the value of his reversion.

99 I can now turn to apply the factors that I have earlier identified to the present case. The first is that the assignment is of six years' rent. It is difficult to say more about the period than that it is significant. Arden LJ makes the G point, which she describes as "significant" (paragraph 55), that the assignment of six rents payable over just five years related to a very short period when compared with the total period of JLP's ownership. I do not think that this is a factor to which much weight should be given. If it were otherwise, odd results would follow. I can illustrate the point by an example. Suppose JLP had a five-year lease of property A and owned the H freehold of property B, and had granted a sublease for three years on the first, and a three-year lease on the second, and then for a lump sum assigned the right to receive the rents to a third party in both cases. Is it to be said that the lump sum is capital in the case of property A, but income in the case of property B? I would suggest not. The owner of the freehold owns the property for an indefinite period. It follows that, if a comparison between

the period to which the rent relates and the duration of the assignor's ownership is significant, it could be said that a sum payable for the assignment of the right to receive payable under a 25-year lease would on that account be properly classifiable as income. I do not believe that this would be correct. A

100 I turn next to the value of the asset assigned. The sum in this case, namely £25·5m-odd, is a substantial sum. Little more needs to be said. Thirdly, there was a diminution to the value of JLP's reversionary interests. That was the unchallenged evidence of Mr Asher. He did not, however, quantify the amount of the diminution. It is reasonable to infer from the size of the lump sum that, at the time of the assignment, the diminution was significant. But the amount of the diminution attributable to the assignment will have reduced with the passage of time. As I have already said, the value must be judged at the date of the assignment and it does not have to be permanent. B C

101 The fourth factor is that the payment in the present case was of a single lump sum. It was not one of a series of recurring payments.

102 Finally, there was a transfer to the bank of the risk of non-payment of the rents. It is, however, true that the effect of the indemnities secured by the bank was that the transfer of risk was somewhat theoretical, since it depended on the insolvency of JL. D

103 In my judgment, the cumulative effect of these considerations is that the payment made by the bank was one of capital. The sum was substantial, it was a single payment for the once and for all disposal by JLP of six years' rents, which resulted in a diminution in the value of its reversionary interests. In reaching my conclusion, I have been influenced by the fact that, if JLP had granted the bank six-year leases at nominal rents, the premiums payable would have been capital payments. For the reasons that I have given, judging the matter from JLP's point of view in a practical common sense way, I consider that the differences between such a transaction and the one which they in fact entered into are not sufficiently significant that they should lead to a different fiscal result. E

104 For these reasons, I think that the judge reached the right conclusion, and I would dismiss this appeal. On the head of charge issue, I agree with the reasoning of Arden LJ. F

SCHIEMANN LJ

105 I am grateful to Arden and Dyson LJJ for their careful analyses of the existing case law. Having no separate contribution of any substance to make I find myself in the unenviable position of having to choose between them. G

106 Like others who are more at home in this field than I am, I have been unable to find a satisfactory conceptual distinction between capital and income which explains all the cases to which Arden and Dyson LJJ have referred.

107 It seems to me reasonably clear that had the present transaction involved the grant of a lease to Rabobank at a nominal rent and with a premium equal to the amount of the proceeds then those proceeds would be regarded as a capital receipt. I understand Arden LJ, in paragraphs 32 to 37, to accept this. H

A 108 What seems on their analyses to be critical is whether the present transaction is for present purposes to be regarded as equivalent to such a transaction. I think it is.

109 Although one talks about the value of property, this conceals the fact that what really falls to be valued is the value of an interest in property. Where Tom, who has an interest in property, grants to Dick out of that interest a lease for a premium, the value of Tom's interest in the land is B undoubtedly diminished in the sense that if Tom sought to sell his interest in the land to Harry, Harry will pay him less than he would have paid had Tom not granted the lease to Dick.

110 In the present case, although JLP has not granted a lease to Rabobank, it seems to me that the value of JLP's interest in the land has been diminished to much the same extent as if it had. Harry would pay C significantly less for any assignment of JLP's interest than he would have paid had the deal with Rabobank not been concluded.

111 Of course it is true that the diminution of the value of Tom's or JLP's interest is matched by the receipt of moneys, but that is true in either case. So that is not a satisfactory ground of distinction between them.

112 We are concerned with the classification of the moneys received by Tom or JLP. For my part I see no reason for distinguishing the two situations D described above. There is room for argument as to whether those moneys should be regarded as capital or as income, but in my judgment the classification should be the same.

113 For the reasons given by Dyson LJ, I consider that the special commissioner and Lightman J did not fall into error in holding that the proceeds should be considered a capital receipt. Since this is the view of the E majority of this court the argument as to head of charge thus does not fall to be resolved.

114 This appeal will therefore be dismissed.

Appeal dismissed with costs.
Permission to appeal refused.

F 3 July 2003. The Appeal Committee of the House of Lords (Lord Nicholls of Birkenhead, Lord Hope of Craighead and Lord Walker of Gestingthorpe) dismissed a petition by the Crown for leave to appeal.

Solicitors: Solicitor of Inland Revenue; Lovells.

HD

G

H

Australasian Tax Teachers' Association Conference
T.C. Beirne School of Law, The University of Queensland, Australia
22 – 24 January 2007

Does the Use of General Anti-Avoidance Rules to Combat Tax Avoidance Breach the Rule of Law?[1]

Rebecca Prebble[2] and John Prebble[3]

Contents

1 Introduction 2

1.1 The rule of law and tax avoidance 2
1.2 General anti-avoidance rules 7
1.3 How do general anti-avoidance rules breach the rule of law? 11
1.4 Why are general anti-avoidance rules especially bad? 13

2 The underlying values of the rule of law 15

2.1 Guidance 15
2.2 Liberty 17
2.3 Human dignity 19
2.4 Effective law 20
2.5 Are general anti-avoidance rules effective? 21

3 Are general anti-avoidance rules nevertheless justified? 27

---

[1] The authors gratefully acknowledge the support of the Henry Lang Fellowship, Institute of Policy Studies, Wellington, towards the writing of this paper.

[2] BA (Hons), LLB (Hons) Victoria University of Wellington, Barrister and Solicitor, Russell McVeagh, Wellington, sometime Intern, Institut za Javne Financijie, Zagreb.

[3] BA, LLB (Hons) Auckland, BCL Oxon, JSD Cornell, Inner Temple, Barrister, Professor and former Dean of Law, Victoria University of Wellington, Senior Fellow, Taxation Law and Policy Research Institute, Monash University, Melbourne, Henry Lang Fellow, Instiitute of Policy Studies, Wellington.

3.1 Problems of income taxation 27
3.2 The importance of certainty..32
3.3 The morality of tax avoidance 34
4 Conclusion 39

1 Introduction

1.1 The rule of law and tax avoidance

"The rule of law" is a compendious term for a number of related values that people generally think good laws should adhere to. Dicey's familiar formulation held that the rule of law requires "the absolute supremacy or predominance of regular law as opposed to the influence of arbitrary power."[4] It is theoretically possible to interpret this condition as requiring merely that there should be laws, as opposed to a series of isolated commands. Nevertheless, theorists writing since Dicey have supplemented Dicey's basic formulation with a number of additional requirements that the basic formulation logically entails if it is to be of any value. The most important additional requirement for present purposes is that the law should be capable of guiding people. In order to guide people, laws must be relatively clear and their application relatively

certain; otherwise people will not know exactly what is permitted and what is forbidden.

That laws should be relatively certain seems at first to be reasonable thing to require. Indeed, governments generally do manage to pass laws that adequately satisfy this condition. However, governments find this criterion difficult to satisfy in the sphere of tax avoidance, specifically when they try to formulate rules to combat that activity.

Tax avoidance is a problem for every country. Avoidance is not evasion. Evasion means lying about one's income, for example, a cash business under-stating its takings. Avoidance is not mitigation. "Mitigation" is not a term of art, but in this article, and generally in the present context, it means reducing one's tax in ways that the statute clearly encourages or permits; for example, taking a deduction for a gift to charity.[5]

Avoidance is between the two. Avoidance means, approximately, contriving artificial transactions to reduce tax that is otherwise payable. This is a description

---

[4] A V Dicey *Introduction to the Study of the Law of the Constitution* (10 ed, Macmillan, London, 1960) 202.

[5] Judith Freedman "Defining Taxpayer Responsibility: In Support of a General Anti-Avoidance Principle [2004] BTR 332, 350.

rather than a definition. Terminology in the area is controversial. Some people deny that we can draw a meaningful distinction between avoidance and mitigation. Some people deny that the word mitigation has any right to exist as a meaningful term in this context.[6]

As a general rule, the law does not demand that people arrange their affairs so that they incur the greatest possible tax liability. When faced with two possible ways in which to organise their money, taxpayers are legitimately entitled to choose the option that requires them to pay the lesser amount of tax. There comes a point, however, where governments begin to think that taxpayers are going too far in their attempts to decrease their tax liability: at this point, taxpayers cease to be engaging in legitimate tax mitigation and begin to be engaging in unacceptable tax avoidance.

Analytic definitions of the point at which tax mitigation becomes tax avoidance are elusive. Lord Denning has said that for an arrangement to constitute tax avoidance, "you must be able to predicate … that [the arrangement] was implemented in that particular way so as to avoid tax."[7] This definition brings us no closer to knowing what constitutes tax avoidance, because all it says is "tax avoidance arrangements are those arrangements that look like tax avoidance arrangements."

---

[6] See, eg, *Miller v CIR* [2001] 3 NZLR 316, 326 Lord Hoffman (PC).
[7] *Newton v FCT* (1958) 98 CLR 1, 8, [1958] 2 All ER 759, 764 (PC).

Nevertheless, the definition highlights the difficulty of exhaustively defining tax avoidance, or, indeed, the difficulty of defining tax avoidance in terms of legal rules at all.

Tax avoidance is perhaps best defined by ostensively, rather than by analysis. That is, the best way to understand tax avoidance is simply to be shown some examples. This exercise proves to be an easier task than exhaustive definition. Tax avoidance transactions tend to have a number of identifiable features, for example, artificiality,[8] lack of business or economic reality,[9] lack of true business risk,[10] and the exploitation of statutory loopholes.[11] Avoidance often involves taxpayers exploiting rules that were designed to reduce unfairness in the tax system[12] or using existing legal structures in enterprising ways.[13]

---

[8] *FCT v Gulland* (1985) 160 CLR 55, 109 Dawson J; *FCT v Spotless Services Ltd* (1996) 186 CLR 404, 425 McHugh J.

[9] *Mangin v CIR* [1971] NZLR 591, 596-598 Lord Donovan (PC), quoting Turner J in the court below.

[10] *Challenge Corporation Ltd v CIR* [1986] 2 NZLR 513, 561 Lord Templeman (PC).

[11] See generally Nabil Orow *General Anti-Avoidance Rules: A Comparative International Analysis* (Jordan Publishing Ltd, Bristol, 2000) 18.

[12] Eg *Challenge Corporation Ltd v CIR* [1986] 2 NZLR 513 (PC) involved a corporate group taking advantage of rules that allowed it to consolidate the affairs of its members and to pay tax only on the net profit. The group tried to minimise tax by buying an outside company that had suffered a loss and subtracting that loss from the profits of the original group.

[13] Eg, *Mangin v CIR* [1971] NZLR 591 (PC) involved an arrangement whereby the taxpayer each year leased the profitable part of his farm, which was a different section each year, to a family trust. The trust would then pay out the income from the section of the land to its beneficiaries, who were the

To help to recognise avoidance, take, for example, *Inland Revenue Commissioners v Bowater Property Developments Ltd*,[14] which the House of Lords decided in 1988. That case involved development land tax, a kind of capital gains tax that applied to land sales if the development value component was more than £50,000. Bowater proposed to sell land, in a transaction potentially caught by the tax, for more than £250,000 to a company called Milton Pipes Ltd.

Instead, Bowater segmented the land into five undivided shares. It sold one share to each of five sibling companies in the Bowater group for £36,000 each. Land in undivided shares looks just like land: there was no subdivisional survey. There were no separate titles. The five Bowater companies owned the land in one title, just as a married couple owns their home in one title. The Bowater companies were a sort of modern marriage with five spouses. These five sales had no effect on the beneficial ownership of the land. Both before and after the sales the ultimate owners were the shareholders in the Bowater group.

---

taxpayer's wife and children. The artificial element in this arrangement was that the part of the farm leased to the trust changed year by year, with the trust always receiving almost all of the farm's income for that year. The result of the arrangement was that each beneficiary received a fraction of the farm's income. The income was therefore taxed at a lower rate than it would have been had it been entirely incurred by the taxpayer.

[14] [1989] AC 398 (HL).

The five companies then sold their undivided shares to Milton Pipes for £50,000 each. That is, each company bought for £36,000 and sold for £50,000, making a profit of £14,000, well under the threshold.

Legally, there were five separate sales from Bowater and five more sales to Milton Pipes. Economically there was one sale from Bowater to Milton Pipes. The House of Lords, however, held that the transactions genuine. Bowater escaped development land tax.

1.2 General anti-avoidance rules

Typically, governments combat avoidance by adding rules to tax legislation that frustrate one kind of avoidance transaction or another. For instance, jurisdictions might allow taxpayer companies to carry losses forward and to set them off against the profits of future years. As an anti-avoidance measure, such jurisdictions tend to require certain minimum continuity of ownership between the loss year and the profit year.[15] Tax statutes are replete with such rules. However, specific anti-avoidance rules cannot combat the more creative forms of tax avoidance that employ transactions that

[15] New Zealand, for example, requires companies to have a minimum continuity of ownership of 49 per cent between loss year and profit year, Income Tax Act 2004, s IE 1(1)(b).

employers cannot predict. Consequently, many tax systems feature general anti-avoidance rules in addition to specific ones.

There is considerable variation in the form that general anti-avoidance rules take in different countries. Nevertheless, the various forms have roughly the same effect, at least in theory. General anti-avoidance rules allow tax authorities to disregard schemes that would otherwise reduce tax liability. The transactions to which they apply are void for tax purposes transactions that it captures. The transaction being void, the tax lies where it falls, though modern general anti-avoidance rules often allow the tax authorities to reconstruct a transaction to reflect the economic reality of the circumstances and tax the taxpayer on the basis of the reconstructed transaction.

An example of a typical general anti-avoidance rule is section 99 of New Zealand's Income Tax Act 1976 (New Zealand's current rule is not so readily quotable because it is disaggregated into several elements,[16] but it has roughly the same meaning and effect). Section 99 relevantly read:

[16] Income Tax Act 2004, section BG 1, incorporating section GB 1 and certain definitions in section OB 1.

Every arrangement made or entered into, whether before or after the commencement of this Act, shall be absolutely void as against the Commissioner for income tax purposes if and to the extent that, directly or indirectly,—

(a) Its purpose or effect is tax avoidance; or

(b) Where it has 2 or more purposes or effects, one of its purposes or effects (not being a merely incidental purpose or effect) is tax avoidance, whether or not any other or others of its purposes or effects relate to, or are referable to, ordinary business or family dealings,—

whether or not any person affected by that arrangement is a party thereto.

Countries that have anti-avoidance rules broadly similar in form to New Zealand's include Australia,[17] Canada,[18] South Africa[19] and Hong Kong.[20] The United States and the United Kingdom do not have statutory general anti-avoidance rules, but they both have judicially developed anti-avoidance rules that can sometimes have roughly the same effect. The United Kingdom common law anti-avoidance doctrine

[17] Income Tax Assessment Act 1936 (Aust.), Part IV A, ss 177A-177G.
[18] Income Tax Act 1988 (Can.), s 245.
[19] Income Tax Act (SA), s 103.
[20] Inland Revenue Ordinance (HK), s 61.

was first propounded by the House of Lords in *W T Ramsey Ltd v IRC.*[21] The United States approach was established by the Supreme Court in *Gregory v Helvering.*[22] Both approaches essentially allow the court to look at a series of transactions and to determine whether the transactions have any economic purpose other than the avoidance of tax. In both countries there have been suggestions that that common law anti-avoidance doctrines are insufficient to combat tax avoidance and should be replaced by general anti-avoidance rules.[23] Civil law countries tend to rely on the "abuse of rights" concept, which forbids the use of rights for improper purposes.[24] The different forms that general anti-avoidance rules take do not affect the associated rule of law issues; the problems and justifications that concern general anti-avoidance rules are equally relevant to all of them.

[21] [1982] AC 300.
[22] (1935) 293 US 465.
[23] See, eg, Judith Freedman "Defining Taxpayer Responsibility: In Support of a General Anti-Avoidance Principle [2004] BTR 332. There is also a bill before the United States House of Representatives that proposes to codify the anti-avoidance doctrine, see H R 1555, "Abusive Tax Shelter and Taxpayer Accountability Act of 2003.
[24] Nabil Orow *General Anti-Avoidance Rules: A Comparative International Analysis*, (Jordan Publishing Ltd, Bristol, 2000) 373.

1.3 How do general anti-avoidance rules breach the rule of law?

The exact content of the rule of law is the focus of an ongoing debate among legal theorists.[25] Nevertheless, as far as certainty is concerned there is close to unanimity: most, presumably all, legal philosophers consider that a law must be relatively certain in order to conform to the principles of the rule of law.[26] It is this requirement of certainly that general anti-avoidance rules offend. Although a number of countries have statutory general anti-avoidance rules, the legislation adds little to the common understanding of what constitutes tax avoidance. By the same token, there is uncertainty in most jurisdictions as to what transactions fall inside and what outside the general anti-avoidance rule.

The uncertainty surrounding tax avoidance stems from the fine line that separates tax avoidance from acceptable tax mitigation. Lord Templeman in *Challenge Corporation Ltd v CIR*[27] considered the two concepts with reference to section 99 of the Income Tax Act 1976, the then New Zealand general anti-avoidance rule:

---

[25] See further, Jeremy Waldron "Is the Rule of Law an Essentially Contested Concept (in Florida)"? (2002) 21 Law and Philosophy 137.
[26] There is abundant support for this proposition, in particular see F A Hayek *The Constitution of Liberty* (Routledge, London, 1960) 144; John Rawls *A Theory of Justice* (The Belknap Press of Harvard University Press, Cambridge (Mass.), 1971), 235.
[27] *Challenge Corporation Ltd v CIR* [1986] 2 NZLR 513, 562.

Income tax is mitigated by a taxpayer who reduces his income or incurs expenditure in circumstances which reduce his assessable income or entitle him to reduction in his tax liability. Section 99 does not apply to tax mitigation because the taxpayer's tax advantage is not derived from an "arrangement" but from the reduction of income which he accepts or the expenditure which he incurs.

Thus when a taxpayer executes a covenant and makes a payment under the covenant he reduces his income. If the covenant exceeds six years and satisfies certain other conditions the reduction in income reduces the assessable income of the taxpayer. The tax advantage results from the payment under the covenant.

...

Section 99 does not apply to tax mitigation where the taxpayer obtains a tax advantage by reducing his income or by incurring expenditure in circumstances in which the taxing statute affords a reduction in tax liability.

Section 99 does apply to tax avoidance. Income tax is avoided and a tax advantage is derived from an arrangement when the taxpayer reduces his liability to tax without involving him in the loss or expenditure which entitles him to that reduction. The taxpayer engaged in tax avoidance does not reduce his income or suffer a loss or incur expenditure but nevertheless obtains a reduction in his liability to tax as if he had.

Although it is generally accepted that general anti-avoidance rules apply to tax avoidance and not to tax mitigation, drawing the line between the two is often

problematic. A literal application of general anti-avoidance rules would catch many legitimate transactions.[28] General anti-avoidance rules therefore mean something more than their bare words.

Probably, there are two meanings of avoidance: first, the ordinary meaning, and secondly the meaning of the words in most statutory general anti-avoidance rules. The second meaning may be a subset of the first. That is, there are perhaps some transactions that people might call avoidance but that are not "avoidance" in the statutory sense. Statutory general anti-avoidance rules do not simply codify ordinary language.

1.4 Why are general anti-avoidance rules especially bad?

The preceding sections of this article have demonstrated that general anti-avoidance rules are vague. However, all legislation is vague to some extent. The most specific rules will always have borderline cases. Why do people single general anti-avoidance rules out as particularly egregious breaches of the rule of law? This paper focuses on general anti-avoidance rules because their vagueness is more systematic than the vagueness of other laws. Drafters of most laws cannot foresee all relevant fact

[28] See, eg, Richardson J in *Challenge Corporation Ltd v CIR* [1986] 2 NZLR 513, 546 (CA) alludes to the somewhat paradoxical consequence situation of a literal interpretation of a general anti-avoidance

situations. As Hart pointed out, all laws admit of "core" situations, where the law will definitely apply, and "penumbra", where it is less certain whether the law will apply.[29] To criticise general anti-avoidance rules because it is not clear whether they apply in some situations appears to subject them to a higher standard than we demand of law in general.

The difference is that general anti-avoidance rules have far larger penumbras than most laws. Arguably, general anti-avoidance rules are nothing but penumbra. The reason why legislators decide that they need general anti-avoidance rules is that situations where general anti-avoidance rules might be needed cannot be defined in advance. If legislators could foresee all varieties of tax avoidance, they would pass specifically targeted rules to frustrate those endeavours. No doubt, most tax policy makers could give examples of the sorts of arrangement that might be caught by general anti-avoidance rules, but these examples would be cases that have been found to constitute avoidance in the past. The fact that general anti-avoidance rules exist at all is evidence that legislators themselves cannot predict what structures taxpayers will eventually contribute.

---

rule being quite obviously not what Parliament intended.

[29] HLA Hart "Positivism and the Separation of Law and Morals" in HLA Hart *Essays in Jurisprudence and Philosophy* (Clarendon Press, Oxford, 1983) 49, 63.

2 The underlying values of the rule of law

2.1 Guidance

The rule of law requires that the law must be certain so that it can provide guidance.[30] Do vagueness and uncertainty offend the rule of law principle that the law should be such that people are able to be guided by it?

This paper examines the deeper values that the requirement of certainty seeks to preserve, and considers whether general anti-avoidance rules truly offend those values. If they do so, are there situations in which the rule of law must give way to countervailing considerations, and is tax avoidance one of those situations? An important factor is the general public tolerance of general anti-avoidance rules. It appears that the rule of law is seen as more important in some areas of law than in others. This paper examines why this is so.

Generally, laws that are as vague as general anti-avoidance rules attract considerable criticism. For example, in 1935 an amendment to the Danzig Penal Code was passed that criminalised acts "deserving of penalty according to the

[30] See, eg, Joseph Raz "The Rule of Law and its Virtue" in Joseph Raz *The Authority of Law: Essays on Law and Morality* (Oxford University Press, Oxford, 1979), 210, 213.

fundamental conceptions of a penal law and sound popular feeling."[31] In an uncomfortable common law echo, the House of Lords in the English case of *Shaw v Director of Public Prosecutions* decided that it had jurisdiction to create new offences in order to punish acts that were contrary to public morals, but that had not previously been held to be illegal.[32]

Both of these examples of uncertainty in the law have been heavily criticised. For example, the Permanent Court of International Justice delivered an opinion condemning the amendment to the Danzig Penal Code.[33] People criticise *Shaw* for similar reasons.[34] Should we be concerned that the reasons that make the Danzig Decree and the decision in Shaw objectionable appear to apply equally to general anti-avoidance rules?

It is difficult to know what effect general anti-avoidance rules have on people's actions. It has been suggested that they act in terrorem, in that people are discouraged from constructing tax avoidance schemes because of the ever-present risk of being

[31] Decree of the Senate of the Free City of Danzig, 29 August 1935, Article 2.
[32] *Shaw v Director of Public Prosecutions* [1962] AC 220 (HL).
[33] Permanent Court of International Justice Advisory Opinion of 4 December 1935
[34] Eg, C C Turpin "Criminal Law – Conspiracy to Corrupt Public Morals" [1961] Cambridge Law Journal 144, 144 -146.

caught by the general anti-avoidance rule.[35] While this consequence is presumably exactly what governments hope for when they resort to general anti-avoidance rules, this effect is not what scholars mean when they argue that the law should be capable of guiding people. However, to demonstrate that general anti-avoidance rules offend the rule of law it is not sufficient simply to show that they do not guide people's actions. To see what is so objectionable about general anti-avoidance rules it is necessary to examine the underlying values of the rule of law, and to reveal why it is important that people should be able to rely on the law to guide them.

2.2 Liberty

The relationship between liberty and laws that can be relied upon is a key part in many theorists' conceptions of the rule of law. For Rawls, people must know exactly what legal rights they can claim because, "If the bases of these claims are unsure, so are the boundaries of men's liberties."[36] An essential part of being free, then, is knowing exactly how free one is. This argument has particular resonance when we look at general anti-avoidance rules. The argument is that general anti-avoidance rules' truly objectionable aspect is that no one really knows how far their reach

[35] See, eg, Michael O'Grady "Acceptable Limits of Tax Planning: A Revenue Perspective" KPMG Tax Conference, Ireland, November 2003, 6.

extends. People are prevented from taking action that might be allowed, the argument continues, because they do not want to take the risk of their action being disallowed.

The argument in the preceding paragraph appears to support the proposition that general anti-avoidance rules offend the rule of law as Rawls sets out that doctrine. But when tax professionals makes this argument they are likely to put it in more specific terms, viz, that the existence of a general anti-avoidance rule has a chilling effect on legitimate tax planning, and that fear of general anti-avoidance rules prevents investors and businesses from utilising effective business structures that appear to be economically sensible.

There may be some truth in this claim, but it is not borne out by reported cases. All cases known to the present writers where the Commissioner has attacked an arrangement using a general anti-avoidance rule involve arrangements that an informed but objective bystander would predicate entail tax avoidance. From another perspective, at meetings of tax professionals one of the writers has frequently asked for examples of transactions or structures that could reasonably be predicated to be legitimate, but that taxpayers have rejected because of fear of a general anti-avoidance rule. Examples have not been forthcoming.

---

[36] John Rawls *A Theory of Justice* (The Belknap Press of Harvard University Press, Cambridge (Mass.), 1971), 235.

Hayek is another philosopher who stresses the connection between the rule of law and liberty, but his conception of liberty is slightly different from that of Rawls. Where Rawls would describe knowledge of the degree of liberty that the law allows as an essential component of liberty itself,[37] Hayek simply sees liberty as the absence of coercion. If people know what the law is in advance, they can choose to put themselves in the position of being subject to it. Subjection to the law is therefore a wilful act.[38] This argument is particularly relevant to general anti-avoidance rules. Since no one knows exactly when general anti-avoidance rules will apply, people who are caught by them have not made a conscious decision to be subject to them, and are therefore coerced.

2.3 Human dignity

For Raz, the criterion that the law should be capable of guiding action is closely linked to human dignity. The law must assume that people are capable of rational thought, and that they therefore want to plan their lives with the knowledge of what the law is.[39] For Raz, this factor is even more important than the rule of law's

[37] Idem.

[38] F A Hayek *The Constitution of Liberty* (Routledge, London, 1960) 144 -145.

[39] Joseph Raz "The Rule of Law and its Virtue" in Joseph Raz *The Authority of Law: Essays on Law and Morality* (Oxford University Press, Oxford, 1979) 210, 221.

connection with freedom.[40] Laws that do not conform to the rule of law, then, are an affront to human dignity, because the law "encourages autonomous action only to frustrate its purpose."[41] Raz might well charge general anti-avoidance rules with such an offence. The detailed formality of tax law encourages people to find ways to circumvent it, but general anti-avoidance rules may frustrate their efforts.

2.4 Effective law and Fuller

It is unlikely that Lon Fuller would disagree with Rawls's argument that the rule of law protects liberty or Raz's proposition that it protects dignity. Fuller, however, focuses his argument on the theory that certain formal criteria of the rule of law must all be sufficiently satisfied in order for law properly so called to exist.[42] Laws must be public, prospective, understandable, non-contradictory, possible to conform to, relatively stable, there must be congruence between how the rules are written down and how they are enforced, and laws must be rules as opposed to ad hoc decisions.[43]

In order to demonstrate how continuous breaches of the rule of law reduce the effectiveness of legal systems, Fuller gives us the example of King Rex. King Rex is

---

[40] Idem.
[41] Id 222.
[42] Lon Fuller *The Morality of Law* (2 ed, New Haven (CT) Yale University Press, 1964) 168.
[43] Idem.

a ruler who tries but fails to make law on eight separate occasions. Each time that Rex attempts to make law, he manages to breach one of these eight criteria. For example, on one occasion Rex publishes a legal code that is so convoluted that no one can understand it and on another occasion he announces that all cases will be decided retrospectively.

Naturally, Rex's subjects are dismayed at their king's disregard for the rule of law, and are annoyed at the way the consequences of that disregard affect them.[44] For present purposes, however, the interesting point is the consequences for Rex. Rex is unable to rule effectively because his rules are incapable of being followed. There is really no point in Rex having laws at all, because his laws do not guide the behaviour of his subjects.[45] However much his subjects might want to obey Rex's laws, they cannot. Fuller's examples show that laws that do not conform to the rule of law can therefore be just as frustrating to law-makers as they are to law-followers.

2.5 Are general anti-avoidance rules effective?

General anti-avoidance rules tend to be counterexamples to Fuller's general theory of effective law. They are frustrating to the citizen, but they are useful to governments.

---

[44] Id 167.
[45] Id 168.

When general anti-avoidance rules work they are undeniably effective, because they allow governments to collect tax that they would otherwise lose. Nevertheless, the experience of some countries with general anti-avoidance rules reveals that they can sometimes be ineffective for reasons very similar to those that plagued King Rex.

For example, when Sir Garfield Barwick was Chief Justice of Australia the Commissioner was seldom successful in litigation where he deployed the general anti-avoidance rule.[46] Barwick CJ felt very strongly that "[i]t is for Parliament to specify, ... with unambiguous clarity, the circumstances which will attract an obligation on the part of the citizen to pay tax."[47] The Chief Justice had little time for the vagueness of the general anti-avoidance rule, and tended to find for the taxpayer even in cases of the most blatant tax avoidance.[48]

Barwick CJ's pro-taxpayer stance reached its apogee in the cases of *Slutzkin v FCT*[49] and *Cridland v FCT*.[50] Slutzkin was a case of dividend-stripping. The taxpayer

---

[46] At the time, Australia's general anti-avoidance rule was contained in s 260 of the Income Tax Assessment Act (Cth) 1936.
[47] *FCT v Westraders Pty Ltd* (1980) 54 ALJR 406, 461 Barwick CJ.
[48] See further, G Lehman, "The income tax judgements of Sir Garfield Barwick: A study in the failure of the new legalism" (1983) 9 Monash ULR 115, 135. Lehman argues Sir Garfield Barwick did not deprive Autralia's general anti-avoidance rule of all effect, because to do so would invite speedy law reform. Barwick CJ allowed the former section 260 to continue to operate where an "antecedent transaction" was involved, see *Mullins v FCT* (1976) 135 CLR 290, 302.
[49] *Slutzkin v FCT* (1977) 140 CLR 314.

was a shareholder in FR Holdings Pty Ltd, a company that was pregnant with profits. Had the company distributed the profits as a dividend they would then have been taxable in the hands of Slutzkin and his fellow shareholders. The same result would have obtained had the shareholders put the company into liquidation and distributed the proceeds.

Instead, the shareholders cashed the company up by liquidating its assets. They then sold their shares to Cadiz Corporation, which was a trader in shares. Cadiz Corporation caused FR Holdings Pty Limited to distribute its retained profits as a dividend. Without its retained profits the company was now worth very little. Cadiz Corporation sold the shares in FR Holdings Pty for nearly all the price that Cadiz Corporation had paid for the shares.

For Slutzkin, the fiscal effect of these transactions was that he sold his shares for a non-taxable capital receipt. Cadiz Corporation Limited, on the other hand, derived a taxable profit from the dividend, but sustained a deductible loss in selling the shares. The loss neatly cancelled the gain from the dividend and left Cadiz Corporation Limited with, in effect, a fee for its trouble. The fee was taxable, but was a very small fraction of the income that Slutzkin and this fellow shareholders had stood to derive from either a profit distribution or a liquidation.

---

[50] *Cridland v FCT* (1977) 140 CLR 330.

Arguing that the only reason that Slutzkin and his fellows sold their shares was to avoid tax on profits that would otherwise have been distributed, the Commissioner submitted that the price of the shares was economically the same thing as a dividend and that the general anti-avoidance rule applied. Barwick CJ rejected this argument, holding that the sale of the shares was "no more than a realisation by them of the benefit of their shareholding in a way which would not attract tax."

Cridland involved a scheme designed to take advantage of a rule that allowed primary producers to average their incomes over a number of years and to pay tax on that average. The rule was intended to make the tax system fairer for people like farmers, whose income often varies considerably from one year to the next. Pursuant to the scheme, Cridland, a university student, bought a share in a unit trust. The trust was a primary producer. Cridland's interest as a beneficiary of the trust was only one dollar a year. The years in which he was a beneficiary straddled his time as a student and also time as a salaried graduate, when his income was much higher. Cridland claimed to average his income as a primary producer. Despite the general anti-avoidance rule, the Barwick court upheld the claim, with Mason J delivering the leading judgment. Both *Slutzkin* and *Cridland* were almost certainly situations where Australia's general anti-avoidance rule should have applied, but Barwick CJ's High Court found in both cases that the taxpayers had not avoided tax.

In response to this judicial attitude, which rendered Australia's general anti-avoidance rule almost useless, the Australian Parliament in 1981 enacted a new type of general anti-avoidance rule that attempts to attain more precision of detail. It is certainly more prolix.[51] In hindsight, Parliament's action was possibly not necessary: following Sir Garfield Barwick's retirement, the High Court was able to re-inject some force into section 260, Australia's then general anti-avoidance rule.[52] The history of how section 260 fared during Sir Garfield Barwick's term as Chief Justice is an interesting example of how the rule of law defects of general anti-avoidance rules can make them ineffective.[53]

It is interesting to note that when general anti-avoidance rules are ineffective, this ineffectiveness is not due primarily to taxpayers being inadequately guided. Rather, when general anti-avoidance rules are ineffective it is because the judiciary do not know what to make of them. To return to general anti-avoidance rules' sinister counterpoint, the amendment to the Danzig Penal Code, it seems that the Nazis had a similar experience. The same rule applied in Germany, as well as in Danzig, but it

[51] Income Tax Assessment Act 1936 (Aust), ss 177A-177G.

[52] See *FCT v Gulland* (1984) 160 CLR 55.

[53] The United Kingdom's experience with a judicially developed anti-avoidance doctrine might be used to illustrate the same point. The doctrine, as developed from its original formulation by Lord Wilberforce in *W T Ramsey Ltd v IRC* [1982] AC 300, 323-326, is so vague that no one seems to be certain whether it even exists. Its application can therefore appear somewhat haphazard, see further, The Right Honorable Lord Walker of Gestinghope "Ramsaqy 25 Years On" (2004) 120 LQR 412.

ultimately led to very few prosecutions in either jurisdiction, because its terms were too vague for the compliant judges of the Nazi era to make much sense of them.[54] As far as tax avoidance is concerned, however, situations where statutory general anti-avoidance rules are ineffective are relative rarities. The majority of jurisdictions that have general anti-avoidance rules find them to be a reasonably effective tool of frustrating tax avoidance.

It is difficult to know what conclusion to draw from the fact that general anti-avoidance rules can be relatively effective. Fuller's argument that laws are more effective when people know what they require certainly seems uncontroversial and likely to be true in most situations. While Fuller does not demand that legal systems must satisfy each of his criteria perfectly in order to conform to the rule of law;[55] it is unlikely that Fuller would approve of the protracted and unapologetic breaches that accompany general anti-avoidance rules.

This point is even clearer if we use Fuller's framework to assess individual laws, as opposed to entire legal systems. A state with some laws that offend Fuller's criteria may still be able to be governed effectively, but, according to Fuller's thesis, an individual rule that continuously breaches many criteria ought not to be effective.

---

[54] Martin Broszat *The Hitler State* (Longman, London 1981) 339.
[55] Lon Fuller *The Morality of Law* (2 ed, New Haven (CT) Yale University Press, 1964) 170.

It is an interesting feature of general anti-avoidance rules that their criteria for effectiveness are almost the exact opposite of the effectiveness criteria of any other law.

3 Are general anti-avoidance rules nevertheless justified?

3.1 Problems of income taxation

The intuitive alternative to a general anti-avoidance rule is a system of very many specific rules that detail exactly what is and is not subject to income tax. Of course, all tax systems already have such specific rules in at least some areas of economic activity, whether or not they also have general anti-avoidance rules. Unfortunately however, the more specific and detailed a system's rules become, the more ways people will find to circumvent those rules.[56] Tax law is unusual in two key respects. First, there are very few other areas of law that people so aggressively try to avoid. Secondly, the nature of tax law means that tax legislation contains a large number of potential loopholes.[57] The result is that in the absence of a general anti-avoidance rule,

---

[56] Judith Freedman "Defining Taxpayer Responsibility: In Support of a General Anti-Avoidance Principle [2004] BTR 332, 346.
[57] For an explanation of why tax law is more susceptible to loopholes than other areas of law, see Ross Parsons "Income Taxation – An Institution in Decay" (1986) 3 Australian Tax Forum 233, as

there is apt to be a great deal of tax avoidance that the government is powerless to stop.

It is tempting to suggest that if legislators cannot frame a tax avoidance rule that conforms to the rule of law they should not have an anti-avoidance rule at all. Governments should just put up with the adverse consequences. However, this suggestion overlooks the fact that tax avoidance is not a problem for governments alone, it is a problem for society generally. Avoidance undermines two key purposes of a tax system. First, the principle of horizontal equity states that people in the same economic position should be taxed at the same rate.[58] Tax avoidance makes horizontal equity difficult to achieve, because successful tax avoidance results in some people being taxed less than others who are in the same economic position. In other words, people who avoid tax are not paying their fair share as measured by their wealth.

Secondly, tax avoidance makes it more difficult for tax systems to be economically neutral. Economic neutrality demands that tax systems should distort the normal workings of the market as little as possible; that is, that people should not make decisions for purely (or even partially) tax reasons. The existence of

---

developed in John Prebble "Income Taxation: a Structure Built on Sand" (2002) 24 Sydney Law Review 303.

[58] See, eg, Richard E Krever "Structure and Policy of Australian Income Taxation" in Richard E Krever (ed) Australian Taxation: Principles and Practice, Longman Cheshire Pty Ltd, Melbourne, 1987, 1, 11.

opportunities for tax avoidance frustrates this goal. To illustrate, consider the case of *Peterson v CIR*,[59] which the Privy Council decided in 2005. *Peterson* was a case in which films were funded principally by non-recourse loans. Pursuant to a scheme, Mr Peterson and others invested in films and deducted their investment from their other income. They were able to depreciate the value of the films to zero over two years.

The promoters of the film told the investors that the cost of the film was (say) $2,000, while in fact it was only (say) $1,000. To fund their investment in the films, Mr Peterson and his co-investors borrowed. The borrowing was in the form of non-recourse loans, that is, loans that were repayable only if the films were successful. Interest was not charged. Loans on such favourable terms naturally attract questions, and indeed it was found as a fact that the money was never borrowed at all.[60] The fact that the extra money from investors was not available did not bother the film's promoters, because they had overstated the cost of the film anyway.

The reason for the overstatement of the cost of the films was the tax saving that it led to. Instead of being able to write off $1,000 over two years, investors were able to write off $2,000, even though they had never actually spent the second $1,000 (and, except on paper, had not even borrowed it). Whether or not the films were successful,

---

[59] *Peterson v CIR* [2005] UKPC 5 (PC). The Privy Council in fact found that New Zealand's general anti-avoidance rule did not apply to this scheme.

the investors would gain a tax advantage. This tax advantage meant that a scheme that would not ordinarily be attractive to investors was in fact attractive.

This situation is a clear example of the tax system creating market distortions: the transactions in *Peterson* were not attractive for their intrinsic merits; they were attractive because of tax advantages. Jurisdictions that have general anti-avoidance rules are able to counteract the effect of this distortion to some extent: to the extent that investors see the tax advantages of a particular scheme as unlikely to stand up to close scrutiny and therefore refrain from investing in it and the market will not be distorted.

The aims of the tax system are related the more general point about the purpose of tax systems. Governments do not tax people only for the sake of amassing wealth. Rather, tax is necessary to keep states functioning. Governments must provide public services such as defence and education. Furthermore, most societies use tax to redistribute wealth to some extent. Tax avoidance reduces the effectiveness of welfare systems,[61] a matter that is particularly important in the light of the public perception (that is probably accurate) that most tax avoidance is perpetrated by the rich or by people who are relatively well off. Though few people have reasoned the issue

---

[60] *Case U32* (2000) 19 NZTC 9,302 [80].
[61] RA McLeod "Tax Avoidance Revisited" (2000) 6 NZJTLP 103, 103.

through to a sufficient depth to put it this way, the wide spread of general anti-avoidance rules, either statutory or judge-made, indicates that most countries consider that the negative results from not having a general anti-avoidance rule outweigh the breaches of the rule of law that general anti-avoidance rules entail.

This balancing exercise reveals much about the nature of the rule of law and its values. Adherence the rule of law can often interfere with a society's other goals. Some philosophers insist that the rule of law must be preserved without compromise.[62] Other writers, such as Raz, stress than the rule of law is only one yardstick against which a legal system may be measured.[63] Just as a society's conformity to the rule of law does not ensure that the society is good, a breach of the rule of law does not make that society bad.[64] Rawls expands on this point, saying that a breach of the rule of law may be "the lesser of two evils."[65] Tax avoidance is a very real evil for society: a breach of the rule of law seems to be a necessary remedy.

---

[62] See, eg, F A Hayek *The Constitution of Liberty* (Routledge, London, 1960).
[63] Joseph Raz "The Rule of Law and its Virtue" in Joseph Raz *The Authority of Law: Essays on Law and Morality* (Oxford University Press, Oxford, 1979), 210, 211.
[64] Even Fuller, who is strongly committed to the rule of law, accepts that isolated breaches do not automatically condemn a legal system. See Lon Fuller *The Morality of Law* (2 ed, New Haven (CT) Yale University Press, 1964),170.
[65] John Rawls *A Theory of Justice* (The Belknap Press of Harvard University Press, Cambridge (Mass.), 1971) 242.

In modern days, at least in democracies that follow a Western model, it is seldom that there is anything sinister about legislators breaching the rule of law. As Fuller observes, laws tend to be most effective when they conform to the rule of law;[66] so governments have a vested interest in making sure their laws conform to its values. In situations where laws offend the rule of law, it will often be the case that the alternative is even less desirable. Tax law is by no means the only situation where the rule of law must be sacrificed to the common good. It is easy to imagine situations where the preservation of human rights or the fulfilment of justice requires a breach of the rule of law. For example, a retrospective law may be necessary to fully compensate for a human rights breach. In such situations, most people would think that breaching the rule of law is justified.

3.2 The importance of certainty

Certainty is clearly an important rule of law value. Usually certainly is important for both the law-follower and the law-maker. Most laws are more effective when people can be certain what they are meant to do or not do. That is, in most cases the rule of law helps to promote effective law. General anti-avoidance rules are therefore an aberration: it is their very vagueness that makes them effective. If they were not

[66] Lon Fuller *The Morality of Law* (2 ed, New Haven (CT) Yale University Press, 1964) 168.

vague, they would not be effective. [67] This characteristic, together with the fundamental problems of tax law plus what many see as the dubious moral standing of tax avoiders, prompts some commentators to argue that certainty is simply an inappropriate value for general anti-avoidance rules to strive for.[68]

Challenge Corporation Ltd v CIR[69] is an example of the negative effect that certainty can have on the utility of an anti-avoidance rule. Challenge Corporation, the taxpayer company, acquired a subsidiary that had suffered heavy losses. Challenge Corporation then purported to set the subsidiary's losses off against its own profits.

At the time, the provisions that allowed intra-group loss consolidation did not require any continuity of shareholding between loss year and profit year. Challenge Corporation had therefore complied with the letter of the law. Without a general anti-avoidance rule, companies in the situation of Challenge Corporation would be able to take deductions despite having suffered no economic loss.

[67] As mentioned in 1.1 above, there are some cases that are definitely tax avoidance (although these cases are mainly ones that have been judicially decided to be tax avoidance); so general anti-avoidance rules' sphere of application is not entirely unknown. Nevertheless, it is true to say that general anti-avoidance rules depend on their vagueness for their effectiveness.
[68] Eg, Judith Freedman "Defining Taxpayer Responsibility: In Support of a General Anti-Avoidance Principle" [2004] BTR 332, 346.
[69] [1986] 2 NZLR 513.

Where the principles of the rule of law negatively influence a law's effectiveness, it is necessary to weigh the consequences of not having the law in question against the possibility that some people will be surprised by the manner in which the law operates. Certainty and related rule of law values are therefore extremely important where criminal sanctions are imposed, but are less important where the issue is tax avoidance.[70]

3.3 The morality of tax avoidance

In the face of such an obvious breach of the rule of law, the fact that so many countries have general anti-avoidance rules seems difficult to account for. The idiosyncrasies of tax law no doubt make general anti-avoidance rules necessary, but it is unlikely that the public tolerance of general anti-avoidance rules is caused by knowledge of these idiosyncrasies. Tax law is extraordinarily complicated, but it is unrealistic to suppose that most people see it as different in kind from other branches of the law. How, then, can we account for the lack of public condemnation of general

[70] While taxpayers are usually extremely annoyed if their tax avoidance schemes are disallowed because of the operation of general anti-avoidance rules, general anti-avoidance rules do not impose *criminal* penalties, although some penalties are involved. It is arguable that it is more important for laws that impose criminal penalties to conform to the rule of law, see, John Rawls *A Theory of Justice* (The Belknap Press of Harvard University Press, Cambridge (Mass.), 1971) 241.

anti-avoidance rules? The explanation may be a perception of tax avoidance as being questionable from a moral perspective.

The moral status of tax avoidance is contentious. There have been a number of cases that hold that since people have the right to arrange their money in such a way as to pay as little tax as possible, even holding that there nothing immoral about tax avoidance.[71] Relying on such decisions, lawyers tend to assume that as a matter of law tax avoidance is morally unimpeachable. However, it is a logical error to say that because tax avoidance is not immoral as a matter of law, it is not immoral in any sense. Whether a certain act is moral must be determined according to ordinary principles of ethics, not by reference to statements in judgments. It is possible that judges who say that there is nothing immoral about tax avoidance are correct, but if that is so it must be because tax avoidance is moral according to ethical principles. As a matter of logic, a judge saying that a particular act is moral as a matter of law cannot determine whether the act is in fact moral.[72]

---

[71] Probably the most famous statement on the morality of tax avoidance comes from Lord Tomlin in *CIR v Duke of Westminster* [1936] AC 1, 19-20 (HL), where his Lordship stated that "every man is entitled if he can to order his affairs so that the tax attaching under the appropriate Act is less than it otherwise would be."

[72] See further, Zoe Prebble and John Prebble "The Morality of Tax Avoidance: Why the Legal difference Between Evasion and Avoidance is Insufficient to Ground a Moral Distinction" Legal Ethics: Professional Ethics and Personal Integrity Conference, University of Auckland, 23-25 June 2006.

According to basic ethical principles, then, what is the moral status of tax avoidance? As a matter of morality untainted by law, people know that they have a duty to pay tax; so seeking to pay less tax than they otherwise might can appear to be shirking that duty.[73] Furthermore, despite the complexity of tax laws, most people have a reasonably clear idea of what the policy of the law would require them to pay. General anti-avoidance rules do not set out to catch individual taxpayers trying earnestly to comply with complex tax laws. Rather, they tend to catch instances of tax planning that is at least relatively aggressive. People who are ultimately caught by general anti-avoidance rules almost always know that they have engaged in something that they would at least concede to be "tax planning" – usually aggressive tax planning - even if they do not expect to be called to account. Taxpayers who engage in tax avoidance schemes are consciously putting other taxpayers at a relative disadvantage and may be criticised on moral grounds.[74]

If the apparent dubious moral status of tax avoidance partially explains the conspicuous lack of public outcry over general anti-avoidance rules, what can we

[73] Nevertheless, the exact amount of tax the each individual should pay is open to debate. It is questionable whether taxpayers who have paid the amount of tax specified by black-letter law can really be shirking a duty. See further, Judith Freedman "Defining Taxpayer Responsibility: In Support of a General Anti-Avoidance Principle" [2004] BTR 332, 337.
[74] See further, The Right Honourable Lord Templeman "Tax and the Taxpayer" 2001 117 LQR 575, 575.

deduce about the relationship between the rule of law and morality? It cannot be correct that people lose their right to rely on the law if they are acting immorally.[75] No one would suggest that the rule of law is unnecessary in the field of criminal law, which typically involves far more obvious immorality than tax avoidance. Possibly the real explanation is that the rule of law itself, as a strict formalist doctrine, inevitably allows people to circumvent the laws that conform to it to some extent. As far as criminal law is concerned, this shortcoming of the rule of law is far outweighed by the benefits that the rule of law offers. In contrast, when it comes to tax avoidance, the benefits to society of legal certainty are outweighed by its detriments.

The argument that the detriments of the rule of law in a particular area outweigh its benefits is nevertheless unsatisfactory. At least, it would not satisfy Hayek, although it might satisfy Rawls or Raz. The merits of the rule of law should not be evaluated on a case-by-case basis, leaving us free to disregard its principles if those principles are inconvenient. Rather, one of the reasons why societies value the rule of law is that it applies *despite* its having a net societal detriment from time to time. Societies commit to adherence to the rule of law for the very reason that there will be instances where it is tempting to allow breaches of it.

[75] But see S Munzer (1982) "A Theory of Retroactive Legislation" 61 Tex L Rev 424. Munzer argues that people have no right to rely on their immoral acts not being retrospectively criminalised.

This argument echoes David Cole's criticism of Richard Posner's *Not a Suicide Pact: The Constitution in a Time of National Emergency.*[76] In his book, Posner argues that the protections offered by the United States' Constitution should be interpreted flexibly in order to allow the government to address the threat of terrorism. Posner argues, for example, that the United States' administration's wiretapping of international phone calls should be considered a "reasonable" search in the context of the threat of terrorism. Cole, however, points out that allowing the provisions of the Constitution to be interpreted more or less strictly according to administrative convenience misses the point of having a constitution in the first place.[77] A constitution like that of the United States, and the rule of law, should be adhered to notwithstanding that doing so is not beneficial to society in every isolated instance. Any kind of cost-benefit analysis is simply inappropriate where the Constitution, and equally the rule of law, is concerned.

The fact that the rule of law is not beneficial in the area of tax law therefore cannot explain the apparent public acceptance of general anti-avoidance rules. What, then, can be the explanation? The most likely answer is that tax avoiders are seen as fundamentally different to criminals. When criminals break the law, they simply break

---

[76] 2006, Oxford University Press.

it; they do not try to find ways of getting around the law so that they have not technically broken it at all. In contrast, there is an entire industry devoted to manipulating tax laws so as to obtain tax advantages without incurring any corresponding economic loss. In the light of this difference, the fact that general anti-avoidance rules are accepted despite their rule of law shortcomings is not surprising.

4 Conclusion

General anti-avoidance rules demonstrate that the rule of law is not an unqualified good. As with all principles, the rule of law can be outweighed by competing considerations. General anti-avoidance rules give an example of what those competing considerations might be. Furthermore, while general anti-avoidance rules themselves are justified, they are useful in showing exactly why we value the rule of law. Most societies with developed legal systems tend not to breach the rule of law very often. As a rare example of a breach, general anti-avoidance rules are a useful reminder of why values such as certainty are important.

---

[77] David Cole, "How to Skip the Constitution" review of Richard Posner *Not a Suicide Pact: The Constitution in a Time of National Emergency*, New York Review of Books, 53:18, November 16 2006.

Ectopia, Formalism, and Anti-Avoidance Rules in Income Tax Law

By John Prebble, Wellington*

Ectopia *Pathol.* Displacement or anomaly of situation or relation . . (OED)

1. Statutory general anti-avoidance rules

Governments attack tax avoidance in a number of ways. Statutory anti-avoidance rules are one means. Such rules come in two forms: specific and general. By 'specific' rules one refers to provisions that are contained in a particular regime within a tax statute, and that are specific to that regime. For example, a regime that permits a company to carry losses forward from one year to the next may include rules that limit this right to cases where the company maintains a certain minimum continuity of shareholding, or where the company continues to carry on the business where the losses were incurred.

General anti-avoidance rules, on the other hand, potentially apply to any kind of transaction that may result in tax avoidance. A typical example is section 99 of the New Zealand Income Tax Act 1976, which reads:

> Every arrangement made or entered into. . . shall be absolutely void as against the Commissioner for income tax purposes if and to the extent that, directly or indirectly, –
> (a) Its purpose or effect is tax avoidance; or
> (b) Where it has two or more purposes or effects, one of its purposes or effects (not being a merely incidental purpose or effect) is tax avoidance, whether or not any other or others of its purposes or effects relate to, or are referable to, ordinary business or family dealings. –
> whether or not any person affected by that arrangement is a party thereto.

Section 99 contains broad definitions of 'arrangement' and 'tax avoidance', the latter including 'avoiding, reducing, or postponing any liability to income tax.' The section empowers the Commissioner of Inland Revenue to reconstruct transactions if, after the section has taken effect and avoided an

* BA, LLB (hons) Auckland; BCL Oxon; JSD Cornell; of the Inner Temple, London; Barrister; Professor of Law in The Victoria University of Wellington, New Zealand. The help of the author's assistant, Mr David Plunkett, in preparing this chapter is gratefully acknowledged, as are discussions with Mr Ian Macduff and Mr B. V. Galvin, which aided in refining the author's ideas.

impugned arrangement, the facts do not disclose a liability on the part of the taxpayer.

With the notable exceptions of the United States of America and the United Kingdom, similar rules are contained in the tax legislation of many common law jurisdictions, including:

- Australia: Income Tax Assessment Act 1936 Part IVA
- Canada: Income Tax Act 1952 section 245
- Hong Kong: Income Tax Ordinance, section 61A
- Malaysia: Income Tax Act 1967 section 140
- Singapore: Income Tax Act section 33.

It is sometimes helpful to think of a statutory general anti-avoidance rule as a kind of rule of interpretation. It is a rule of interpretation that borrows from both the English style of strict interpretation and the American style of free interpretation. These two approaches to statutory interpretation are discussed by Atiyah and Summers in *Form and Substance in Anglo-American Law*. Atiyah and Summers consider the case where the draftsman has made an error, with the result that the plain meaning of the words of a statute fails to give effect to the intent of the legislature. They point out that the American lawyer would argue that the statute should be interpreted according to its substance and policy, in order to correct the error, whereas the English lawyer, while not denying that injustice will sometimes ensue, will hew to the plain meaning of the words. For the English lawyer, to start from the assumption that the draftsman is less than competent is likely to lead, in the long run, to more error.[1]

As mentioned, a general anti-avoidance rule may be seen as incorporating elements of both approaches. To the American it says: seek out the substance and do justice. To the English lawyer it says: the draftsman has indeed captured the intention of the legislature in the specific rules of the income tax legislation, but there is a more fundamental intent that must prevail. This fundamental intent is that people must not be permitted to avoid tax by exploiting the strict forms of the legislation.

2. The United States of America

The United States Internal Revenue Code contains no general anti-avoidance provision. This vacuity does not leave a deficiency in the weapons of the United States revenue authorities of the same magnitude that a similar omis-

[1] P. S. Atiyah and R. S. Summers, Form and Substance in Anglo-American Law (Oxford 1987) 106.

sion might leave in some other common law countries. The reason is that, by the standards of other common law judges, United States judges take a relatively anti-formalist approach to law in general and to tax law in particular.

American judges are more willing to hold documents to be shams or to take a substance over form approach to tax cases than are their counterparts in other common law jurisdictions. Consequently, American judges are willing to ignore the strict legal form of documents and the strict rights of the parties to them, and to levy tax on the basis of the substance of the transaction in question. That is, when American courts act to frustrate tax avoidance transactions, or to tax a transaction according to its economic effect, they rely on their inherent powers as courts and not on statutory authority to that effect.[2]

3. The United Kingdom

The United Kingdom lacks a general statutory anti-avoidance rule in its income tax legislation, but, at the same time, its judges take a relatively formal approach to law. One result is that transactions that would in other jurisdictions be brushed aside as artificial and avoidance-driven are apt to be tolerated in the United Kingdom. A celebrated example is *Inland Revenue Commissioners v Duke of Westminster,*[3] where the House of Lords upheld an arrangement whereby the Duke contrived to pay his gardener out of before tax income. By covenant, the Duke diverted income from investments direct to the gardener. The alternative was for the Duke to derive the income, to pay tax on it, and to remunerate the gardener out of after tax income. The House of Lords held that the Duke should be taxed according to the legal form of the transactions, and not according to their economic substance.

A more modern example is *Inland Revenue Commissioners v Bowater Property Developments Ltd,*[4] which concerned development land tax. At the time, this tax applied to the realised development value of land above a threshold of £50,000. The taxpayer was concerned about the vulnerability to the tax of a certain plot of land. The taxpayer formed a number of subsidiaries and divided the land among them and itself as tenants in common. That is, the land was not divided into separate titles. It was common ground between the taxpayer and the tax inspector that the transfers were made solely to avoid development land tax. Nevertheless, the taxpayer argued that there was a separate value threshold for development land tax purposes in respect of the beneficial interest in land held by each company. The House of Lords agreed.

---

[2] The United States form-over-substance doctrines as applied in tax cases are usefully summarised in B. J. Arnold and J. R. Wilson, 'The General Anti-Avoidance Rule – Part I' (1988) 36 *Canadian Tax Journal* 880, 881.

[3] [1936] AC 1; sub nom *Duke of Westminster v IRC* 19 TC 490.

[4] [1989] AC 398 HL.

The formalism of the English approach is these days somewhat bluntened by a judge-made rule known as the 'fiscal nullity' doctrine. The extent and limits of this doctrine are still in the process of being worked out but, broadly speaking, the doctrine takes a preordained series of transactions that contains a step that is inserted only in order to avoid tax and treats the series as a single transaction for tax purposes, in effect ignoring the avoidance step. Generally, for the doctrine to apply it must be practically certain at the time of one transaction that the others would follow.[5] The fact that the taxpayer in the *Bowater* case successfully navigated the shoals of the fiscal nullity doctrine gives some indication of its limitations as a weapon in the hands of the revenue authorities.

A second recent development in the United Kingdom is for the courts to examine very carefully the actual legal effect of transactions, or of a series of transactions, in order to determine precisely the legal rights and obligations to which they give rise. That is, the courts are not bound by the labels that parties give to their transactions, nor by the form of the transactions, if name and form do not correspond with legal effect. This approach is not new, but it is nowadays taken further than formerly. For example, and over-simplifying somewhat, in *Ensign Tankers (Leasing) Ltd v Stokes*[6] the House of Lords analysed a transaction that was in form a non-recourse loan, and came to the conclusion that in legal effect it was a joint venture. Their Lordships held that the transaction should be treated for tax purposes as a joint venture and not as a non-recourse loan.

4. The rule of law

The development of the fiscal nullity doctrine has attracted considerable criticism in the United Kingdom. The burden of the criticism is that the doctrine substitutes uncertainty and judge-made law for the rule of law as enacted by Parliament.[7] General anti-avoidance rules contained in statutes, such as section 99 of the New Zealand Income Tax Act, attract similar criticism.[8] A fundamental problem is that whether a general anti-avoidance rule is statutory or judge-made its scope is uncertain. As Lord Templeman said in a New Zealand

[5] For a useful summary of the doctrine and an examination of the leading cases see J. Tiley (ed.), *Butterworths UK Tax Guide 1992 - 1993*, 6 - 23.

[6] [1992] 2 All ER 275; [1992] 2 WLR 469 HL.

[7] See, e.g., Revenue Law Committee of The Law Society, *Tax Law in the Melting Pot* (The Law Society 1985) 19ff., and 'The Discourse of Resistance' in Doreen McBarnet and Christopher Whelan, 'The Elusive Spirit of the Law: Formalism and the Struggle for Legal Control' (1991) 54 *Modern Law Review* 848, 856ff.

[8] See, e.g., G. J. Harley, 'Legislative and Judicial Anti-Avoidance Rules', in Asian-Pacific Tax and Research Centre, Singapore and Institute of Policy Studies, Wellington, *Anti-Avoidance and Tax Treaty Policies in the Asian-Pacific Region* (1989) 18 - 28.

appeal to the Privy Council, *Challenge Corp Ltd v Commissioner of Inland Revenue:*

> In the first place, even the [Commissioner] concedes that section 99, albeit expressed in the widest possible terms, has to be read subject to some limitation as regards transactions permitted or authorised by other legislative provisions if it is not to produce results that are absurd.[9]

The reason for this limitation is that some transactions that clearly have a tax-limiting effect are equally clearly envisaged, even promoted, by Parliament. Therefore, they should not be frustrated by an anti-avoidance rule. For example, many jurisdictions encourage savings by affording tax preferences to life insurance premiums or superannuation contributions. A taxpayer who takes advantage of such a preference is clearly liable neither to a statutory nor to a judge-made general anti-avoidance rule, but where does one draw the line? The answer is that it is not possible to draw the line. The law is worked out on a case-by-case basis, with all the uncertainty that this entails, whether the anti-avoidance rule in question is statutory or judge-made.[10]

The disfavour in which lawyers tend to hold statutory general anti-avoidance rules in particular is further explained by factors identified by Atiyah and Summers in *Form and Substance in Anglo-American Law*. Typically, statutes draw clear lines that are easily discoverable and known to all, whereas precedent is by nature less definite and more flexible. Further, statute law is embodied in a fixed, canonical form. It is the paradigmatic type of formal law, and is susceptible to a high degree of interpretive formality.[11] Statutory general anti-avoidance rules not only fail to comply with the paradigm; they unashamedly reject these typical properties of statute law. General anti-avoidance rules direct the courts to adopt a substance-over-form approach, and avowedly introduce uncertainty into a regime (income tax law) that is in most countries highly formalistic and already complex enough. The reaction of the lawyer, who is trained to respect the merits of form, is not surprising.

---

[9] [1986] 2 New Zealand Law Reports 555, 564 (PC).

[10] But see Lon Fuller, who contended in 'American Legal Realism' (1934) 82 *University of Pennsylvania Law Review* 429, 434 - 435 that formalistic rules create more uncertainty than open-textured, informal rules. Fuller's argument was that, in a formal system, judges first decide a case on its merits and then torture the case into compliance with a more-or-less appropriate formal rule, whereas a judge who is governed by open-textured directives can rely on policy justifications, which enable a more logical system of rules to be built up. One could regard the development of the United Kingdom fiscal nullity doctrine as a vindication of Fuller, in that the doctrine is arguably a systematic refusal by English judges to observe the formal rules that governed them. See further, Jason Scott Johnston, 'Uncertainty, Chaos, and the Torts Process: an Economic Analysis of Legal Form' (1991) 76 *Cornell Law Review*, 341.

[11] P. S. Atiyah and R. S. Summers, *Form and Substance in Anglo-American Law* (Oxford 1987) 96 - 97. See also R. S. Summers, 'The Formal Character of Law', (1992) *Cambridge Law Journal* 242, 248, for a list of reasons for, or benefits that flow from, the formal aspect of law.

5. The demands of a taxation system

The existence of general anti-avoidance rules is contrary to the values of certainty, form, and the primacy of the rule of law. But there is a competing value: the need to ensure that a country's tax system works effectively, and is not subverted by ingenious tax planning. In most of the countries mentioned in this chapter this competing value has prevailed, and general anti-avoidance rules have been enacted as legislation.

To put the matter another way, the simple explanation for the desirability of a general anti-avoidance rule is that the forms of business transaction and of investment are infinite and cannot all be predicted by the makers of tax policy. Consequently, governments cannot predict, and cannot comprehensively guard against, the measures that people may adopt to avoid tax. Formal, certain, focused anti-avoidance rules will inevitably have loopholes, which will inevitably be exploited. A general rule acts as a back-stop to narrow, regime-specific rules.

There must be some sympathy for any government that is faced with the problem of minimising tax avoidance and trying to achieve some sort of equity between taxpayers in respect of the effective tax rates that they suffer. Scholars and practitioners in the tax area are very familiar with the way in which expertly-advised citizens are able to minimise their tax burdens, to the detriment of the taxpaying public in general.[12]

Be that as it may, a general anti-avoidance rule *does* erode the the rule of law in the manner that has been described here. In other areas, we are unwilling to see the rule of law give way to practical requirements of the state. One has to think only of the procedural safeguards that protect people charged with criminal offences, or the principles developed in the common law whereby the courts are able and willing to review administrative actions of the government.

The existence of a general anti-avoidance rule of uncertain scope is defended by reference to practicalities, but is that enough? More fundamentally, is the argument even morally or intellectually respectable? A telling consideration is that the argument is an example of 'ends justify means', which is an unfortunate basis for an erosion of the rule of law.

This reflection leads to the questions: is there some other argument that might justify the adoption of general anti-avoidance rules? More specifically, is there some characteristic of tax law that relevantly distinguishes it from the law in general and that justifies the resort to uncertainty that is inherent in a general anti-avoidance rule? The thesis of this chapter is that the answer is yes.

[12] See, e.g. Doreen McBarnet and Christopher Whelan, 'The Elusive Spirit of the Law: Formalism and the Struggle for Legal Control' (1991) 54 *Modern Law Review* 848.

Tax law, or, at least, income tax law, has several characteristics that distinguish it from most other areas of law. These characteristics justify a substance over form approach in at least some areas of tax law.

6. Fundamental flaws in the foundations of income tax law

It is basic to income tax law that there must be a concept of income. Taxpayers, the tax authorities, and the courts, must be able to say that a particular receipt or profit is income, and therefore taxable, or that it is not. Further, it is basic to income tax law that some income items are taxable and that some are not. For example, states do not attempt to tax foreign income derived by foreigners. For a tax system to work fairly and efficiently the rules that determine which receipts are income, and, of those, which are taxable, should be logical and consistent with one another.

This basic requirement for rules that can identify receipts that are income and that can determine which income items are taxable can never be perfectly satisfied, even in theory. The problem is that income, in the sense that the term is used in income tax law, is an artificial creation of that law. It is not something that one meets in nature or in the world of reality. This proposition may be illustrated by reference to the connection of income on the one hand with space and time on the other hand.

7. Space

The public law of a state operates within the frontiers and upon the citizens or residents of that state. Thus, with a few exceptions in respect of crimes like piracy, states do not attempt to use their criminal laws to regulate behaviour that takes place beyond their frontiers. Nor will a court of one country attempt to review the actions of an official of another country. In private law, the rules are different. One state will, for example, enforce a contract made somewhere else, or enforce the inheritance rules of another state.

Income tax law is part of a state's public law. No state purports to tax *any* income. All states have rules to determine what income they will tax, and what income they will leave alone. Most states tax income that arises within their borders. Many states also tax income of residents that arises abroad. Some tax income of citizens that arises abroad. But states do not purport to tax the foreign source income of foreigners.

The statement of the rule contains within itself the statement of the problem: for purposes of income tax law it is often necessary to determine whether an item of income has a foreign source or a domestic source. Here lies the difficulty. If the analogy is not over-strained, income tax law is part of public law,

but income itself has more the features of something that is related to private law.

Take interest. The earning of interest is something that citizens arrange for themselves, by virtue of rules of private law, and something that they could arrange perfectly well (indeed, more easily) if there were no national frontiers. But what is the source of interest that is paid by a resident of France to a resident of Germany, in respect of a loan that is secured over land in Canada, where the contract was negotiated and signed in England, where the physical payment of the interest takes place by automatic transfer between two Hong Kong banks, and the loan monies are used as circulating capital in a business that is carried on in the United States of America?

None of the jurisdictions mentioned would have much difficulty in enforcing a judgment for the interest, and they would have little more difficulty in adjudicating a dispute on a more fundamental question about whether the interest was payable in the first place. But when it comes to the question of *taxing* the interest problems of principle arise for any country that taxes income on the basis of source.

The problem is that intangible, abstract things like interest cannot have a source in the sense that, for example, a river has a source somewhere in the physical world. Nevertheless, tax law has no alternative but to treat the source of interest as if it were a physical place that can be located in space, geographically. The reason is that national frontiers are a feature of the physical world: they divide it into physical parts; and the existence of national frontiers forces us to identify the source of interest as if it were a physical phenomenon. That is not possible, if only because money is not a physical phenomenon. (Bank notes and other tokens of money exist physically, but only primitive tax systems would answer questions of derivation of interest by reference to the time or place at which bank notes pass from one person to another. Such systems are particularly vulnerable to avoidance.)

The same difficulties arise in respect of business profits. What is the source of the profit derived by an airline that carries a passenger between Bombay and Bangkok, being one leg of a journey from Frankfurt to Sydney? Does it make any difference where the passenger's ticket was bought, or where the airline is incorporated? What is the source of profit from a contract to manufacture goods in Italy and to export them to the United States? Does it make any difference where the contract was concluded, or where property in the goods passes?

Again, there is no logical answer to these questions. The profits are derived not *because of* but *in spite of* or *ignoring* national frontiers. As in the case of interest, the frontiers are an artificial, man-made matrix that are imposed ab extra. Be that as it may, any country that taxes income on the basis of its

source must have laws that identify the source of income. For the reasons mentioned, such laws are inherently flawed in the same sense that a rule that purports to measure, say, anger, in terms of centimetres or kilograms is flawed. That is, when one takes into account the manner in which income tax rules must cope with problems of space or geography it becomes clear that there are inherent flaws in the very foundations of the discipline.

These flaws open the way to tax avoidance. There are an infinite number of ways in which people can exploit the spatial limitations of the concept of income in order to avoid tax, but one example will suffice here. Suppose a taxpayer has one million dollars with which to build a factory. The obvious approach is simply to pay a builder to erect the factory, and to move in. An alternative is to use the million dollars to capitalise a company in a tax haven (that is, a country that charges no income tax) and to borrow the capital back at interest, and then to pay the builder, thus creating an artificial deduction for the benefit of the taxpayer. Whether this scheme is worthwhile depends on whether, or on how much, tax is charged on the interest that flows outward to the company in the tax haven. But it is not uncommon for tax planners to be able to achieve a low or zero tax rate on outward flowing interest by one technique or another.

8. Time

The problem of time is similar in kind to the problem of space. The income of a typical business is a stream that continues from year to year without reference to the calendar. But an income tax regime cannot operate in this manner. For there to be income tax, income must not only be identified, but it must be divided into manageable portions. The government cannot wait until a business winds down in order to see what income the business has derived over its life. Rather, the life of the business must be divided into periods and the income of each period must be assessed and taxed.

The period chosen is invariably 12 months, because the year is a duration that people are comfortable with, and that is used for other purposes, including the reporting of profits to owners. Because the year is used in so many contexts its use for taxation is hardly a cause for thought, let alone for comment. But governments could just as well have opted for longer or shorter periods had they so chosen, without making much difference to the rules of income taxation.

Be that as it may, once a government has decided to measure income by reference to time, no matter what length of time, there must immediately be whole codes of rules that are calculated to ensure that people do not contrive to defer their income to a later period, or to accelerate their expenses to an earlier period. Take the taxpayer who lends $ 500 for two years at 12 per cent

interest, payable monthly, a total of $ 120. Assume that the 24 months correspond to two fiscal years of the tax regime to which the lender is subject.

From an economic point of view, the monthly payments received by the taxpayer are the same as receipt of all the interest on day one (with an actuarially-calculated decrease in the $ 120 to account for the time value of money) or of receipt of the total sum on day 731, the first day of year three (with a corresponding increase in the $ 120). But in a simple system that assesses income to tax annually on a receipts basis, there will be a considerable difference between the three situations as far as tax consequences go.

Where the interest is all received on day one the taxpayer will pay all his tax in year one. Where the interest is paid monthly tax will be divided equally between years one and two. Where the interest is paid on day 731 tax will be paid in year three. The taxpayer's position is worse in the first case because his fiscal liability is accelerated, and better in the third case because it is deferred. Because of the time value of money, tax deferred is tax saved.

For this sort of reason, most tax systems have rules of greater or lesser sophistication that endeavour to force taxpayers to report their income on an accruals basis that spreads profits evenly into each tax year. As time goes by and as governments learn from one another, tax systems become increasingly mature and such rules become increasingly effective. But the very essence of the rules is that they must treat transactions as if they were different from their factual form, and as if the legal rights of the parties were different from the rights that are set out in their contracts. That is, for example, they must treat the third case in the example given as if interest was spread over 24 months and not in fact received on day 731, and the actual sums deemed to have been derived for tax purposes must be adjusted to allow for the time value of money.

There are other problems with time. Take the farmer who, like most farmers, is prey to the cycles of the market. She enjoys five good, profitable, years, followed by five bad years, where she incurs losses. Suppose that her net result over the ten years is to break even. In most tax systems the farmer will lose, because the tax that she has paid in the first five years will not be refunded in the second five years. To cope with this kind of problem some tax regimes allow farmers, and others whose income is notably lumpy, to pay tax on an average basis rather than on the basis of actual profits.

9. Capital

One timing problem dwarfs all others. This is the distinction between capital and income, a distinction that virtually all regimes that purport to tax 'income' attempt to make. The problem is that there is no clear line that distinguishes

capital from income and, even worse, it is relatively easy for taxpayers to convert income into capital, and thus to avoid tax. Of course, most tax regimes have rules that are designed to frustrate this kind of activity, but they are not always successful.

The difficulty, and ultimate illogicality, of distinguishing capital from income is illustrated by considering the question of business premises. If a business leases its premises it will deduct the rent that it pays, along with other expenses, from its receipts in calculating its taxable profits. If the business buys its premises it will not take a deduction for the price, because the price is a capital item, though the law may make the taxpayer an allowance in calculating profits to compensate for depreciation of the premises from year to year. If one form of expenses on premises is capital, and another form is income, what is the difference? To over-simplify, capital expenses relate to things that constitute enduring benefits for the taxpayer, and income items relate to benefits that are more temporary. But business premises perform precisely the same function whether they are rented or purchased.

The lack of a true distinction between capital items and income items may be illustrated more graphically by considering some simple transactions that are postulated to take place in the absence of a tax regime. Suppose a retailer, a sole trader, sells up her business. To her, there is no difference between the cash received for the various components: premises, fittings, and trading stock. In this context, the distinction between capital and income is meaningless for her purposes.

Alternatively, a retailer may consider improving her sales both by increasing her stock and by renovating her shop. For tax purposes she must treat the expenses differently. Admittedly, for the purposes of calculating her profits on an annual basis (if she is so disposed even in the absence of a tax system) she must also categorise the expenses differently. But if it is a question of determining profits over the long haul (say 20 years) both items are simply expenses that have been repeated a number of times, though one more often than the other.

This similarity of function between capital items and income items is fundamental to the ability of taxpayers to change income into capital. Take, for example, a portfolio of shares that is pregnant with profits, held by a private investor. The shareholder who wants to realise and to use the profits has two options. First, she can wait for a dividend to be declared. But a dividend is an income item and is taxable. Alternatively, she can sell the shares, and the proceeds will reflect both the capital value of the shares and the potential dividend. If the taxpayer is a private investor, the proceeds will be capital in her hands and not subject to income tax.

Of course, the benefits to the shareholder of this characterisation of the proceeds may be limited in jurisdictions where there is a capital gains tax, and the

income tax rules of the jurisdiction may also have something to say on the matter. But these considerations do not detract from the fundamental point that, by its very nature, an income tax mandates a distinction between capital and income, a distinction that does not have a basis in economic reality.

10. The ectopia of tax law

Preceding sections of this chapter have explained that it is fundamental to an income tax that there should be a definition of 'income' and a delimitation of the geographical scope of the tax, and that both of these rules should be logical and internally consistent. At the same time, it has been argued that the need to fit the concept of income into the dimensions of space and time means that these fundamental requirements can never be perfected.[13] This paradox is a manifestation, or perhaps a consequence, of a more general phenomenon that may be called the 'ectopia' of income tax law.

This term is used by the author to refer to a characteristic that distinguishes tax law from most other areas of law: the fact that tax law is in a sense detached and dislocated from its subject matter. The phenomenon is particularly apparent in rules that relate to the taxation of business profits. Tax law clearly *refers* to business profits, but it does not describe the phenomena by which business profits arise, and its objective is not to regulate profits or the methods by which they are won.

Compare the way that tax law works with, say, criminal law. Criminal law defines certain actions as crimes and provides penalties for them. That is, criminal law is closely related to its subject matter, which is criminal activity. Similarly, laws that regulate business activity describe certain activities that are unlawful, or prescribe procedures for certain kinds of transaction. Antitrust laws, for example, proscribe behaviour that is characterised by the law as monopolistic, and Statutes of Frauds require writing and a signature before certain kinds of contract are able to be enforced.

All this is not surprising; in fact, were it not for the contrast that obtains in respect of tax law the fact that law is usually closely related to its subject matter would hardly be worth mentioning. After all, if one sets out to write a law about crime it is to be expected that the result will indeed relate closely to criminal activity. There is a symbiotic relationship between criminal law and crime, just as there is a symbiotic relationship between most law and its subject matter. To use Weinrib's words, '... the distinctive rationality of law is immanent to the legal material on which it operates'.[14]

---

[13] The problems that space, time, and capital pose to the legislator who composes income tax law are explored further in J. Prebble, *Why is Tax Law Incomprehensible?* (1993) IX Victoria University of Wellington Inaugural Addresses (New Series).

Tax law is different. In the business area we start with business activity, an activity that exists independently of tax law. Furthermore, the aspect of business in which tax law is most interested, profit, also exists independently of tax law. Tax law is imposed from above on phenomena that already exist, business and profit. Moreover, far from attempting to describe, mould, or influence business and profit, tax law is most efficient when it is neutral between different forms of business association or investment, between different kinds of transaction, or between different industries.

The purpose of tax law is to raise taxes, not to regulate or to promote business. In order to promote administrative efficiency and to minimise compliance costs, and, indeed, to increase the chances of taxes' working at all, governments ensure that tax legislation is as much in sympathy with the forms of business that the legislation is designed to tax as is possible. But in the end, as has been mentioned, tax law is an alien matrix that is imposed upon existing economic activity that, in the nature of things, it cannot fit exactly. As has been explained, because tax law must classify and delimit income by reference to space and to time, and because income is inherently not apt to be subjected to these measures, there is a dislocation, or ectopia, between tax law and its subject matter.

The ectopic nature of tax law may not be unique, but finding analogies in other areas of law is difficult.[15] One that occurs to the author is the imposition by a conqueror of its family law on a subject people. Suppose that a nation whose family law is based on monogamy conquers a people for whom polyan-

[14] Ernest J. Weinrib, 'Legal Formalism' (1988) 97 *Yale Law Journal* 949, 953, and see id., 955 and passim, and at 974: 'Intelligibility involves the interpretation of content and its immanent form. One achieves a complete understanding when the form is exhibited and the content is seen as adequate to it. If the elements initially identified have the truly fundamental significance that legal experience claims for them, they will be constituents in the distinctive unity that makes the juridical relationship what it is. The immediate understanding of legal experience is only provisional until form becomes explicit. Then juridical intelligibility emerges from a mutually reinforcing movement between form and content: form is the organising idea latent in the content of a sophisticated legal culture, and the ultimate test for legal content is its adequacy to the form it expresses.' A major thesis of this chapter is that income tax law cannot pass this 'ultimate test' because of fundamental conceptual inconsistencies in its makeup.

[15] It is emphasised that by 'ectopic' law one refers to law that, logically, can relate only imperfectly to its subject matter. From ectopic law should be distinguished law that, on examination, appears to be incoherent and lacking in principle. The cause of such incoherence and lack of principle may be ectopia. But it may alternatively be that although there is no general agreement on fundamental principle, law makers and scholars treat the law as if it were a coherent matrix. Application of the law may show that there is in fact inconsistency and incoherence, but it is at least conceptually possible that the law might be re-written in a manner that is coherent. Income tax law, on the other hand, can never be wholly coherent because of the impossibility of rationalising rules about time and space. An example of an examination of an area of law said to be incoherent is John W. Van Doren, 'Private Property: A Study in Incoherence', (1986) 63 *University of Detroit Law Review* 683.

dry is the norm, and imposes its family law on the conquered. How are the courts to deal with a dispute between a wife and one of her husbands? Even more difficult, how will the courts cope with a dispute between two husbands of one wife? In short, there will be an ectopia between the law and its subject matter. No doubt partly for that reason, conquerors much more commonly impose their public or commercial law on the conquered than their family law.

The ectopia of tax law is never as apparent as the hypothetical ectopia of family law described in the previous paragraph. One reason is that parliaments try as best they can to fit tax laws to the structures and transactions of the businesses to which the laws apply. This process masks the problem and, be it said, over large areas of tax law solves it. An analogy is a circle (tax law) that covers a square (business) with sides of the same length as the circle's diameter. Coverage is effective over most of the square, but the corners escape.

11. Ectopia and formality in tax law

Not surprisingly, the greatest opportunities for tax avoidance tend to occur where the ectopia of tax law is most apparent – at the corners of the square, as it were. Thus, diverting income to a foreign country to take advantage of the factor of space, or deferring the recognition of income, or converting income to capital, in the latter cases to take advantage of the factor of time, describe in general terms a large proportion of all the techniques of avoidance that are available to the taxpayer. In response, large parts of the income tax legislation of most countries are given over to specific rules that are calculated to frustrate these methods of avoidance.

Typically, these specific anti-avoidance rules are highly formal, which is another feature of tax law. One reason for this formality is the ectopia of tax law; there is a fundamental discontinuity between tax law and its subject matter. Consequently, tax law must rely on form rather than on substance. The substance that it deals with (business) is not the substance of its raison d'être (raising money).

The thesis of the previous paragraph is more familiar when expressed in another manner: for the courts, there is no policy content in tax law. The most frequently quoted judicial statement to this effect is probably this passage from Mr Justice Rowlatt's judgment in *Cape Brandy Syndicate v Inland Revenue Commissioners*:[16]

[16] [1921] 1 King's Bench 64, 71. The lack of policy content in tax law is highlighted by a comparison of tax law with other areas of law. In other areas, it is not so hard for a judge to escape formalism because she generally has some criterion other than the plain language of the statute to which to refer; that is, the policy of the statute can be made out, and the judge can refer to principle, either general or specific. Cf Frederick Schnauer, 'Formalism' (1988) 97 *Yale Law Journal*, 509, 518 - 519.

.. in a taxing Act one has to look merely at what is clearly said. There is no room for any intendment. There is no equity about a tax. There is no presumption as to a tax. Nothing is to be read in, nothing implied. One can only look fairly at the language used.

So deeply ingrained is this habit of thought that courts will sometimes decline to interpret tax legislation purposively even when directed to do so by statute. For example Section 5 (j) of the New Zealand Acts Interpretation Act 1924 provides that:

'Every Act ... shall be deemed remedial, whether its immediate purport is to direct the doing of anything Parliament deems to be for the public good, or to prevent or punish the doing of anything it deems contrary to the public good, and shall accordingly receive such fair, large, and liberal construction and interpretation as will best ensure the attainment of the object of the Act ... according to its true intent, meaning, and spirit...'

In *Commissioner of Inland Revenue v International Importing Ltd*,[17] Mr Justice Turner, President of the Court of Appeal, said that the approach enjoined by section 5 (j):

'... is normally of little material assistance in the construction of revenue statutes. The 'object of the Act' which the section designates as a key to questions of statutory construction is often only too clearly the collection of funds to swell the general revenues of the State; and Courts of construction have consistently declined to read implications into such statutes to catch a taxpayer who in his business dealings has relied upon the text of the statute, by some extension of the wording accepting the notion of a moral duty to pay a 'proper' amount of tax. The taxing provision is read as prescribing the tax for which its text plainly provides, no more and no less.'

12. An illusory face of formalism

A modern tax statute is a complex construction of intricate, interlocking, formal rules. Superficially, it appears, as it were, to be aware of its mission and to be carrying out that mission in a businesslike and purposeful manner. There is some truth in that impression. As Summers points out, the highly formal rules of the tax system do in fact function relatively smoothly most of the time, and do in fact enable the state to raise huge sums of money in revenue.[18] To use again the metaphor employed earlier, in the area of the square that is covered by the circle tax laws operate with some certainty. Elsewhere there is an elusive substance of, say, profit, the essence of which cannot adequately be captured by rules that relate to a matter that is external to the profit, that is, rules that are part of an income tax regime that attempts to locate the profit in

[17] [1972] New Zealand Law Reports 1095, 1096, CA.

[18] R. S. Summers, 'The Formal Character of Law', (1992) *Cambridge Law Journal* 242, 250.

time and space. In short, the surface appearance of an apparently robust, integrated regime is belied by foundations that are elusive and shifting.

The result can only be a failure of the income tax system to capture gains that its authors intended that it should capture (or would have intended, had they thought of the matter), as lacunae are discovered between formal rules.

This situation lends weight to, and may even justify, an argument for a general anti-avoidance rule. A general anti-avoidance rule is in essence anti-formalist and substantive. But its role is to extend the coverage of an income tax regime to areas that, almost by definition, an income tax regime cannot reach, being areas that elude the law because of the impossibility of ever fitting the naturally occurring phenomena of business and profits into a regime that can operate only within the boundaries of artificial constructs: national frontiers and the calendar.

To put the conclusion another way, the concept of profit is not amenable to definition in terms of rules that fix the profit of a particular taxpayer in time and space. And yet any income tax regime requires such rules to be used for just that purpose. Consequently, such rules can only be formal and, in the end, arbitrary. Being arbitrary, they tend to leave substantive lacunae that escape their grasp. These areas of substance can be reached only by a rule that itself is anti-formalist, namely a general anti-avoidance rule.

[To digress briefly, the fact that a jurisdiction adopts a general anti-avoidance rule, and the arguments that may justify that adoption, do not inevitably lead to the the rule's being applied in the liberal manner that its authors no doubt intend. McBarnet and Whelan usefully list and discuss a number of techniques whereby open-ended anti-avoidance and similar anti-formalist rules sometimes become refined and formalised (the techniques include, for example, self-restraint and the issuance of guidelines by government authorities, and the narrowing of precedents by the courts).[19] The emasculation of a previous form of the Australian anti-avoidance rule (section 260 of the Income Tax Assessment Act 1936) at the hands of the Australian High Court is notorious.[20] Not so dramatically, in the late 1980s the United Kingdom fiscal nullity doctrine was limited and formalised by the courts[21] to such effect that nowadays it is more often called simply the 'New Approach' (ordi-

[19] Doreen McBarnet and Christopher Whelan, 'The Elusive Spirit of the Law: Formalism and the Struggle for Legal Control' (1991) 54 *Modern Law Review* 848, 860ff.

[20] E.g., *Cridland v Commissioner of Taxation* (Cth) (1977) 140 Commonwealth Law Reports 330, 77 ATC 4538; *Slutzkin v Commissioner of Taxation* (Cth) 140 Commonwealth Law Reports 314, 77 ATC 4076; *Mullens Investments Pty Ltd v Commissioner of Taxation* (Cth) (1976) 135 Commonwealth Law Reports 290, 76 ATC 4288. See generally, G. Lehmann, 'Judicial and Statutory Restrictions on Tax Avoidance', in Richard E. Krever (ed.) *Australian Taxation, Principles and Practice* (Melbourne 1987) 295 - 313.

[21] See generally J. Tiley, ed, *Butterworths UK Tax Guide 1992 - 1993*, 6 - 23. In the early 1980s the New Approach appeared to be likely to develop into a true substance-

narily with the capital letters appropriate to a proper name), the label 'fiscal nullity' perhaps being thought to promise more to the revenue authorities than the courts deliver. But the *fate* of a general anti-avoidance rule, as opposed to the reason or justification for its existence, is not germane to the thesis of this chapter.]

It is an interesting irony that the reason for the relative formality of tax regimes is that there are substantive requirements of a tax system that cannot be framed in terms of coherent rules. These requirements cannot, therefore, reasonably be left to the judges, because judges are expected to systematise their decisions in a logical, consistent manner. The rules must therefore be framed as statutes, which can be arbitrary and inconsistent but nevertheless good law.[22] A further irony is that income tax law, that most formal of areas of law, is at bottom incapable of satisfying the indispensable theme of formalism: the internal intelligibility of law.[23] As will have been appreciated, this fundamental lack of internal intelligibility, this substantive incoherence that is an inherent characteristic of income tax law, is different in kind from the general indeterminacy of law that is described by members of the critical legal studies school, and is not an example of that indeterminacy (to express no view on whether those who adhere to the indeterminacy doctrine are correct).[24]

As has been mentioned, in many regimes the formal rules of income tax law are reinforced by a highly informal general anti-avoidance rule, a rule that is so informal that it might even be said to lack the 'essence of law'.[25] But this lack of formality is explained by the fact that income tax law in general lacks something that might also be called an essence of the law: a symbiotic relationship between the law and its subject matter.

over-form doctrine, but it became formalised in, particularly, *Craven v White* [1989] AC 398, 62 TC 1, [1988] 3 All ER 495 HL.

[22] Cf P. S. Atiyah and R. S. Summers, *Form and Substance in Anglo-American Law* (Oxford 1987) 96 - 97.

[23] See Ernest J. Weinrib, 'Legal Formalism' (1988) 97 *Yale Law Journal* 949, 952.

[24] For a discussion and evaluation of the idea of indeterminacy, see Ken Kress, 'Legal Indeterminacy', (1989) 77 *California Law Review*, 283, and see generally, Kelman, *A Guide to Critical Legal Studies* (Cambridge, Mass, 1987).

[25] R. S. Summers, 'The Formal Character of Law', (1992) *Cambridge Law Journal* 242, 260.