



Achieving Better Regulation of Services

Conference
Proceedings

*Australian National University
Canberra, 26-27 June 2000*

© Commonwealth of Australia 2000

ISBN 1 74037 027 9

This work is subject to copyright. Apart from any use as permitted under the *Copyright Act 1968*, the work may be reproduced in whole or in part for study or training purposes, subject to the inclusion of an acknowledgment of the source. Reproduction for commercial use or sale requires prior written permission from AusInfo. Requests and inquiries concerning reproduction and rights should be addressed to the Manager, Legislative Services, AusInfo, GPO Box 1920, Canberra, ACT, 2601.

Inquiries:

Media and Publications
Productivity Commission
Locked Bag 2
Collins Street East Post Office
Melbourne Vic 8003

Tel: (03) 9653 2244
Fax: (03) 9653 2303
Email: maps@pc.gov.au

An appropriate citation for this paper is:

Productivity Commission and Australian National University 2000, *Achieving Better Regulation of Services*, Conference Proceedings, AusInfo, Canberra, November.

The Productivity Commission

The Productivity Commission, an independent Commonwealth agency, is the Government's principal review and advisory body on microeconomic policy and regulation. It conducts public inquiries and research into a broad range of economic and social issues affecting the welfare of Australians.

The Commission's independence is underpinned by an Act of Parliament. Its processes and outputs are open to public scrutiny and are driven by concern for the wellbeing of the community as a whole.

Information on the Productivity Commission, its publications and its current work program can be found on the World Wide Web at www.pc.gov.au or by contacting Media and Publications on (03) 9653 2244.

Foreword

In June 2000, the Productivity Commission and the Australian National University jointly hosted a conference in Canberra on the topic *Achieving Better Regulation of Services*. The conference drew together Australian and international academics, industry participants and policy makers with an interest in regulatory issues.

The purpose of the conference was to learn from experience around the world in how best to design and implement services regulation to ensure good outcomes, both nationally and internationally. The topics covered ranged from the rationales and objectives of regulation in areas such as competition, consumer protection and social policy, to issues to do with measuring the impacts of regulation (particularly on trade in services).

The papers and discussions were of high quality. This conference volume has been designed to capture the diversity of views, by reproducing the papers presented at the conference and the discussants' comments on those papers as well as providing summaries of the general discussion in each session. It is hoped that this will stimulate further research and help guide policy development in this important area.

We are grateful to everyone who participated in the conference, especially those who prepared papers. Special thanks go to the participants from overseas: Aaditya Mattoo and Will Martin (from the World Bank), Pierré Sauve (Harvard University, now at the OECD), Patrick Messerlin (Institut d'Etudes Politiques de Paris) and Roger Kerr (New Zealand Business Roundtable). We are also grateful to the conference organisers — Greg McGuire, Philippa Dee and Christopher Findlay — and to those who assisted in running the conference and assembling this volume — Owen Gabbitas, Duc Nguyen-Hong and Samantha Doove. Dave Cobau and Marilyn Popp provided valuable administrative support.

Gary Banks
Chairman
Productivity Commission

Christopher Findlay
Professor of Economics
Asia Pacific School of Economics and
Management
Australian National University

Contents

Foreword	III
Abbreviations and explanations	X

Part A Introduction

1	Introduction <i>Gary Banks</i>	1
---	-----------------------------------	---

Part B Why regulate services?

2	Introduction to the regulation of services <i>Christopher Findlay</i>	9
	Discussant <i>Patrick Messerlin</i>	25
	General discussion	28
3	The principles of good regulation <i>Paul Coghlan</i>	31
	Discussant <i>Aaditya Mattoo</i>	54
	General discussion	58

Part C Regulating for competition

4	Access: what, where and how <i>Stephen King</i>	63
	Discussant <i>Ed Willett</i>	94
	General discussion	101

5	Regulating access to telecommunications networks <i>Robert Albon</i>	103
	Discussant <i>Rod Shogren</i>	122
	General discussion	125
6	Improving the regulation of international air services <i>Geraldine Gentle</i>	127
	Discussant <i>Christopher Findlay</i>	152
	General discussion	155
Part D Regulating for consumer confidence		
7	Lessons from the Asian Crisis <i>Jeff Carmichael</i>	159
	Discussant <i>Tom Valentine</i>	170
	General discussion	173
8	Principles for regulating the professions <i>Neil Byron, Lisa Gropp and Kim Gusberti</i>	177
	Discussant <i>Jeffrey Keddie</i>	198
	General discussion	202
	Response <i>Neil Byron</i>	204
9	Regulating for the safe and effective delivery of pharmacy services <i>David Greig</i>	205
	Discussant <i>Helen Lapsley</i>	219
	General discussion	222

Part E Regulating to attain social goals

10	Is market-based regulation the best way to regulate community services?	
	The case for market-based regulation <i>Alan Moran</i>	227
	The case against market-based regulation <i>Robert Fitzgerald AM</i>	238
	Response <i>Alan Moran</i>	246
	Response <i>Robert Fitzgerald AM</i>	247
	General discussion	248
11	Regulating gambling: a ‘market friendly’ approach to the social impacts <i>Gary Banks</i>	251
	Discussant <i>David Johnson</i>	278
	General discussion	285
12	Regulating culture: has it ‘Gone with the Wind’? <i>Patrick Messerlin</i>	287
	Discussant <i>Pierre Sauvé</i>	319
	General discussion	321
13	Towards multilateral rules on trade and culture: protective regulation or efficient protection? <i>Pierre Sauvé and Karsten Steinfatt</i>	323
	Discussant <i>Patrick Messerlin</i>	347
	General discussion	349

Part F	Measuring and modelling the impact of restrictions on trade in services	
14	‘Sins of commission and omission’: measuring regulatory impediments to trade in services <i>Tony Warren and Christopher Findlay</i>	355
	Discussant <i>Alison Burrows</i>	370
	Discussant <i>Will Martin</i>	374
	General discussion	376
15	Modelling the liberalisation of services <i>Philippa Dee, Kevin Hanslow and Tien Phamduc</i>	379
	Discussant <i>Kym Anderson</i>	411
	Discussant <i>Will Martin</i>	416
	General discussion	418
Part G	Achieving better regulation of services	
16	Approaches to improving regulation: regulatory competition or regulatory harmonisation? A forum	423
	General discussion	440
Appendixes		
A	Conference participants	445
B	Conference program	451

BOXES

3.1	Regulatory tiers	32
3.2	Current examples where Commonwealth regulation of services is proposed or under consideration	35
3.3	The regulatory spectrum	38
3.4	Elements of a Regulation Impact Statement (RIS)	46
3.5	Comments on the RIS for an access regime for postal services	48
5.1	Access pricing regimes	110
11.1	What is problem gambling?	261
11.2	'Ballpark' estimates of the benefits and costs of gambling	263
13.1	Clashing over the meaning of culture	328
13.2	Domestic support measures used in cultural industries	338

FIGURES

3.1	Notional relationships between risk and regulatory forms	37
3.2	Stylised policy development process	47
7.1	Bank assets as a percentage of financial sector assets	164
10.1	Fees charged by childcare centres	231
11.1	The nature of gambling regulations	253
11.2	The importance of gambling taxes	255
11.3	Prevalence and spending shares of problem gamblers	261
11.4	The impact of lifting caps in NSW hotels	266
11.5	Consumer protection measures	267
11.6	The distribution of social benefits from gambling	282
14.1	Comparing marginal cost and marginal benefit of an intervention	375
15.1	Direct and indirect input requirements per unit of final demand	394

TABLES

11.1	Measured consumer benefits, social costs and net impacts of gambling, by mode of gambling	264
11.2	How often do you withdraw money from an ATM at a venue when you play the poker machines?	271
12.1	Funding sources of subsidies to EC audiovisual services	300
12.2	Use of subsidies and subsidisation rates in EC audiovisual services	301
12.3	French audiovisual policy: 'Au nom de la culture'	306
15.1	Classifying barriers to trade in banking and telecommunications services	388

15.2	Tax equivalents of post-Uruguay barriers to trade and investment	390
15.3	Partial equilibrium effects on selected regions of removing global barriers to trade in services, by sector	399
15.4	Partial and general equilibrium effects on sectoral output in selected regions of removing global barriers to trade in services	403
15.5	Welfare effects of full multilateral liberalisation of services trade	404

Abbreviations and explanations

Abbreviations

ACCC	Australian Competition and Consumer Commission
ACCI	Australian Chamber of Commerce and Industry
ACP	Australian Council of Professions
AFTA	ASEAN Free Trade Area
AGMMA	Australian Gaming Machine Manufacturer's Association
AGPS	Australian Government Publishing Service
AIHW	Australian Institute of Health and Welfare
ANU	Australian National University
APEC	Asia Pacific Economic Cooperation
APO	Australian Post Office
APRA	Australian Prudential Regulation Authority
ASA	Air service arrangements
ASEAN	Association of South East Asian Nations
ATM	Automated telling machine
AUSTEL	Australian Telecommunications Authority
BSAP	Building Surveyors and Allied Professionals Accreditation Board Inc
BBC	British Broadcasting Corporation
BEP	Business entry point
BIE	Bureau of Industry Economics (Australia)
BTCE	Bureau of Transport and Communications Economics (Australia)
CAN	Customer access network
CAP	Common Agricultural Policy (European Union)

CAPs	Common Audiovisual Policies (European Union)
CBA	Cost-benefit analysis
CEO	Chief executive officer
CER	Closer Economic Relations (Australia and New Zealand)
CES	Constant elasticity of substitution
CET	Constant elasticity of transformation
CNC	Centre National de la Cinématographie (France)
CNRS	Centre National de la Recherche Scientifique (France)
COAG	Council of Australian Governments
CPA	Competition Principles Agreement (Australia)
CPE	Customer premises equipment
CPI	Consumer price index
CSA	Conseil Supérieur de l'Audiovisuel (France)
CSC	Committee on Specific Commitments (WTO)
DAIC	Directly attributable incremental cost
DEWRSB	Department of Employment, Workplace Relations and Small Business (Australia)
DFAT	Department of Foreign Affairs and Trade (Australia)
DIAS	Development Industry Accreditation Services Ltd
DRTB	Department of Transport and Regional Development (Australia)
EC	European Community
ECPR	Efficient component pricing rule
ERA	Effective rate of assistance
EU	European Union
EV	Equivalent variation
FACS	Department of Family and Community Services (Australia)
FDC	Fully distributed cost
FDI	Foreign direct investment
FLTSLRIC	Forward looking total service long-run incremental cost
FTAP	GTAP model with foreign direct investment

GATS	General Agreement on Trade in Services (WTO)
GATT	General Agreement on Tariffs and Trade (WTO)
GBE	Government business enterprises
GDP	Gross domestic product
GTAP	Global Trade Analysis Project (Purdue University)
HMO	Health Maintenance Organization (United States)
IAC	Industries Assistance Commission (Australia)
IAP	Individual action plans
IATA	International Air Transport Association
IC	Industry Commission (Australia)
ICAO	International Civil Aviation Organization
IPART	Independent Pricing and Regulatory Tribunal (NSW)
ITU	International Telecommunications Union
LEN	Local exchange network
LRMC	Long-run marginal cost
M-ECPR	Market determined efficient component pricing rule
MFN	Most favoured nation (WTO)
NAFTA	North American Free Trade Association
NCA	National Commission of Audit (Australia)
NCC	National Competition Council (Australia)
NCP	National Competition Policy (Australia)
NGBT	Negotiating Group on Basic Telecommunications (WTO)
NSW	New South Wales
NTB	Non-tariff barriers (WTO)
OECD	Organisation for Economic Co-operation and Development
ORG	Office of the Regulator General (Victoria)
ORR	Office of Regulation Review (Australia)
OTC	Overseas Telecommunications Commission (Australia)
PBS	Pharmaceutical Benefits Scheme (Australia)
PC	Productivity Commission (Australia)

PGA	Pharmacy Guild of Australia
PMG	Postmaster-General (Australia)
PSA	Prices Surveillance Authority (Australia) (chapter 5)
PSA	Pharmaceutical Society of Australia (chapter 7)
RAS	An iterative scaling technique named after Richard A. Stone
RIS	Regulation impact statement
ROR	Rate-of-return
ROSA	Review of Structural Arrangements (Australia)
SAGIT	(Cultural Industries) Sectoral Advisory Group on International Trade (Canada)
SP	Service providers
SPS	Sanitary and Phytosanitary (WTO)
SRMC	Short-run marginal cost
STD	Subscriber trunk dialling
TAB	Totalizator Agency Board
TAF	Telecommunications Access Forum (Australia)
TPA	<i>Trade Practices Act 1974</i> (Cwlth) (Australia)
TRIPS	(Agreement on) Trade-related Aspects of Intellectual Property (WTO)
TSLRIC	Total service long-run incremental cost
TV	Television
TWF	Television Without Frontiers (European Union)
UK	United Kingdom
UNCTAD	United Nations Conference on Trade and Development
UNESCO	United Nations Educational, Scientific and Cultural Organization
US	United States
VANS	Value-added network services
VCR	Video cassette recording
WPDR	Working Party on Domestic Regulation (WTO)
WTO	World Trade Organization

PART A

INTRODUCTION

1 Introduction

Gary Banks

Productivity Commission

Service industries collectively loom very large in all countries, especially the more developed economies. Australia is no exception, with service industries accounting for over three-quarters of gross domestic product (GDP), four-fifths of employment and one-fifth of (recorded) exports.

International trade in services is also important — estimated to exceed US\$1.3 trillion worldwide or approximately one quarter of all trade (WTO 1999) — and it continues to grow strongly.

Service sector performance is critical to community living standards, both through its contribution to income growth and because of the nature of the activities themselves, many of which have important social and cultural dimensions. The sector also contains some of the most dynamic industries, with technological advances in information and communications and transport services being key drivers of globalisation. These important industries are also highly regulated, often in ways that generate substantial costs as well as benefits, both domestically and internationally.

So the sector is clearly important, and poses some complex challenges for regulators. But why have a conference? Policy analysts and advisers as well as academic researchers have learnt a lot about the costs of regulation and how to meet social and economic objectives in better ways. But much of this learning is recent and somewhat dispersed — and some of it is prospective. It is therefore useful to swap ideas, pool knowledge and look for some common lessons and themes, and to identify where further work is most needed.

Australia's reform experience

As the host country for the conference, it is worth noting that there is much to learn from Australia's own experience. We appear to have been more dogged and systematic than many other countries in our reform efforts, perhaps because

Australia started out so badly with a highly protected and cosseted economy (long justified under the oxymoronic banner ‘protection all round’).

I am not suggesting that Australia’s approach has necessarily involved an *optimal* sequencing of reforms — a vexed and complex issue — but there has been some political-economic logic in the progression of the reform program in this country.

It began with the progressive liberalisation of border restrictions in goods and capital markets. This unleashed competitive pressures on a range of trade exposed industries, which looked at ways to reduce their costs and improve their productivity, leading in turn to a further round of reforms in key input markets such as public utility services and other economic infrastructure and, of course, labour markets, the regulation of which had historic links to the regulation of trade.

More recently, the reform process has been extended to general government administration, social services and the systematic review of a wide range of regulation with anticompetitive or other cost impacts on business. Opening the Australian economy has also driven the need for governments to curb budgetary excesses and demonstrate good financial performance. Reform has involved changes in governance of (previously) government-owned business enterprises (GBEs), and exposure to competition, to enhance incentives to be cost-conscious and productive.

The link between domestic and international reform is a core issue in services, as it is in other industries. The scope of this conference accordingly ranges from the key domestic issues to some international dimensions, including the institutions involved.

Domestic and international dimensions

Australia stands out among many OECD countries in its unilateral approach to liberalising merchandise trade, which it has done for the domestic gains, largely independently of the actions of other countries. It has then sought credit for such liberalisation by undertaking to reduce bound tariff rates in multilateral negotiations under the GATT. Our approach has been similar in the case of barriers to international trade in services, where domestic liberalisation — again based on its perceived domestic benefits — has generally preceded any international commitments.

For other countries, the GATT has played a more active or initiating role in the liberalisation of merchandise trade. And this is being extended to services, through the General Agreement on Trade in Services (GATS) in the World Trade Organization (WTO). But progress is very uneven and there are a number of

deficiencies in the GATS framework compared to the GATT, including the ‘opting-in’ and selective approach to coverage, and the level of obligations. This not only reflects the complexity of the regulatory framework for services, but also the special political sensitivities of exposing many services — which are seen as being in the realm of domestic policy — to international tradeoffs and foreign competition. This reaction was apparent in the aborted Multilateral Agreement on Investment, which can now be seen as a precursor to the more visible agitation at the WTO Ministerial Conference in Seattle.

While the reform process in service industries has involved a certain amount of deregulation, in most cases the decision is not *whether* to regulate, but rather the *kind* of regulation that is most ‘appropriate’. In other words, regulation needs to meet legitimate objectives or rationales in a cost-effective way.

Good regulatory outcomes demand good policy-making processes. These need to encompass three key questions.

- What is the nature of the social or economic problem requiring government attention?
- What are the available options for dealing with it (including *non-regulatory* options)?
- What is the most cost-effective option or solution?

In addressing these questions, it is important to have an understanding not only of the direct or localised effects of regulatory options, but also their economywide and international dimensions. And just as analysis of the domestic effects is critical to achieving a beneficial outcome from international negotiations, assessment of reform options nationally needs to take account of relevant international treaties and institutional frameworks. International rules can also constitute a politically useful discipline on regulatory discretion in the national interest.

Governments clearly need good information and analysis if they are to choose wisely. And this information needs to be effectively communicated to the public. This is true of all regulation, but it is particularly important for many service activities, given their complex social and cultural as well as economic dimensions. It is also important given the considerable community resistance to economic frameworks and market mechanisms, particularly in areas like health, education and community services.

Notwithstanding these perceptions, the fact is that economic incentives matter in all areas of the economy and society more broadly. The challenge for policymakers and regulators is to find ways of improving incentives such that governments’ objectives on the community’s behalf are more cost-effectively satisfied. That the market can

be harnessed in the pursuit of equity as well as efficiency is not adequately understood: otherwise economic rationalism would not have become such a pejorative in Australian public discourse!

That said, it is also true that policy actions in the name of reform have not always met the tests of good process *or* good economics. And even where a priori tests are met, one can rarely be certain about all the likely effects of a particular regulatory intervention, especially in areas of rapid change.

A topical example is the uncertain future course of the digital revolution and its implications for today's regulatory decisions in telecommunications and broadcasting. As the Productivity Commission noted in its recent broadcasting report:

In such an environment, regulation must be flexible enough to deal with uncertainty and change. It should not advantage some technologies or media producers at the expense of others. (PC 2000, p. 6)

The Commission's contribution

The Commission and its predecessors have played a leading role in the analysis, development and review of regulation in the service sector. Not as policymaker or regulator, but as an independent source of advice and public information, bringing into account the long-term effects on the wider community.

In line with the reform progression outlined earlier, most of this contribution has occurred in more recent years. The earlier focus was on measuring and reforming distortions in the merchandise sector (where there was a lot to do).

The Commission's incursion into the service sector began with reviews of economic infrastructure services in the public domain, including government-owned business enterprises in transport (rail, shipping, air), energy, water and communications. More recently, there has been an increasing involvement in services with strong social dimensions; examples being charities, public housing, occupational health and safety and workers compensation. Some of the most recent inquiries have had a mix of economic, social *and* cultural aspects — the broadcasting and gambling inquiries being cases in point.

The Commission has recently also embarked on some reviews of established reform programs and regulatory frameworks in the service sector, such as in rail and telecommunications, and there is a forthcoming review of the provisions in the Trade Practices Act (Part IIIA) governing third party access to certain essential infrastructure.

These are proving to be very challenging areas for achieving good regulatory design and implementation at the national level. It is more challenging still to attempt to negotiate and coordinate liberalisation internationally.

The latter has a particular requirement to identify and measure barriers to trade in services, based on a good understanding of the different mechanisms through which trade in services occurs.

The Productivity Commission's predecessors made a substantial contribution to quantifying impediments to trade in *goods*, drawing on the pioneering work on effective rates of protection by Max Corden at the Australian National University (ANU). It is therefore fitting that those institutional linkages between the Commission and the ANU have been maintained in a major joint exercise to measure barriers to trade in services. The Commission is also pleased to have the opportunity to co-host this conference with the ANU.

Structure of this conference volume

The papers presented in Part B of this conference volume consider the rationales for regulating services, the objectives and characteristics of good and bad policy design, including when and where regulation is appropriate. Part C explores regulation designed to promote competition by focusing on the key network industries — rail, telecommunications and international air services. Yet promoting competition is only one of a number of broad objectives of service regulation. The papers presented in Part D look at measures to promote consumer confidence with the professions — banking and finance, and architectural and pharmacy services. Regulation has also traditionally sought to achieve social goals — part E focuses on regulation in the fields of community services, gambling and culture. It includes two debates on the best way to achieve social goals. The authors in Part F consider how to measure the effects of regulation and the consequences of international regulatory reform. The volume concludes with a forum drawing together many of the issues spanning the conference.

References

- PC (Productivity Commission) 2000, *Broadcasting*, Report no. 11, AusInfo, Canberra.
- WTO (World Trade Organization) 1999, *International Trade Statistics, 1999*, WTO, Geneva.

PART B

WHY REGULATE SERVICES?

2 Introduction to the regulation of services

Christopher Findlay

Australian National University

2.1 Introduction

Why regulate services? Most regulation of the service sector seeks to address market failures. However, regulation is only one of a number of possible remedies available to government, each with their own set of costs and benefits. Therefore, a rigorous assessment weighing up the costs and benefits of the various possible policy responses will help to ensure that the objectives are met in the most efficient manner possible.

Regulation should improve the welfare of society as a whole. Policymakers should consider the effects of regulation on all members of society. Yet private interest pressures can and do influence regulatory arrangements, either by encouraging government to choose the wrong remedy or apply the appropriate remedy inefficiently. As a result, the policy response actually applied can differ from that which is socially most efficient.

Inefficient regulation does not optimally address the underlying objectives. In particular, inappropriate regulatory choices can adversely affect trade and investment. A common approach to examining these effects is to model them as shocks to the marginal and fixed costs of foreign suppliers compared to local suppliers. Such an approach has important implications, especially on the prices of imports and domestic services, and on the balance of trade. The former effects are more predictable. Yet such regulatory differences, especially in the area of meeting domestic standards in the importing country, may actually advantage foreign suppliers by helping them to gain domestic acceptance.

Researchers and policymakers need to be able to measure and assess the effects of regulation to better manage the process of regulatory reform. Different regulatory

instruments have substantially different effects. But instruments can be broadly characterised according to the effects that they have.

While the ultimate objective of regulation should be to maximise domestic social welfare, international trade and investment considerations should not be overlooked, as they have extremely important welfare repercussions. Therefore, policymakers should take these effects into account when assessing regulation.

There are numerous strategies for improving domestic regulation. International agreements — including the General Agreement on Trade in Services (GATS) — have an important role to play. They enable the benchmarking of regulation, lower trade costs and avoid political problems. The use of horizontal disciplines within international agreements facilitates reform, not only because they lower the cost of negotiation, but also because they ameliorate private interest issues in the reform process.

2.2 Objectives in regulating services

The public interest

The public interest case for service sector regulation is to solve market failure problems. These problems can be divided into four groups:

- lack of information;
- natural monopoly;
- externalities; and
- social objectives (eg income distribution or service quality).

All these issues arise in the services sector. In some cases, the nature of the provision of services and the direct contact involved between producer and consumer generate significant problems and risks, as well as consequences that are not easily reversible for consumers who lack information about the skill of their service provider. Many service sectors are examples of network industries where control of core parts of the network amounts to a significant barrier to entry. Provision of various services can have significant positive external effects, such as those associated with health or education. Services can also generate negative effects. Some people might say, for example, that direct contact between foreign service providers and consumers in the provision of education might have negative cultural impacts. Another example is the negative externalities associated with tourism.

However, while a market failure problem might be apparent, it is not necessarily the case that government intervention in the market is appropriate. This depends on a ranking of the instruments available and an assessment of the costs and benefits of their application.

In many cases, market mechanisms will emerge or market processes will develop to solve the problem (eg eventually entry occurs to solve a market power problem or new ways of providing information to consumers are developed). Simply proposing to do something about market failures can invite a rent-seeking response from incumbents already in the market, thereby making intervention more likely, and increasing the risk of regulatory capture and even worse outcomes occurring. Therefore, in considering the choice and design of regulatory instruments, it is important to retain the option of leaving the problem to the market to resolve.

Usually, there is more than one instrument that might be used to solve each type of problem.¹ The literature concludes that the appropriate instrument to choose is the one that best addresses the problem directly (Vousden 1990, pp. 30–8). The tariff, for example, can be argued to be a second best instrument for achieving the public policy objectives often ascribed to it, as there are more efficient instruments available for dealing with income distribution objectives. Similarly, in the service sector, a ban on the foreign supply of some service, supposedly to deal with asymmetric information problems, is less efficient than alternatives that keep the market open, but which provide information to consumers.

This approach of ranking instruments assumes that governments efficiently implement the best instrument. But whenever governments intervene, there is also a risk of government failure, because of many factors, including insufficient information or deliberate action in response to lobbying (see chapter 14 in this volume). The risk of government failure is also high in sectors where there is rapid technological change. Even though, in principle, a particular instrument might be the most appropriate, it may be less so once the difficulties of implementation are identified. The impact of pressure from private interest groups is discussed below.

The choice of instrument may not be straightforward, since there may be a number of instruments that can directly achieve a given objective.² For instance, apart from

¹ Another important principle is that, where there are a number of policy objectives, an equal number of policy instruments are required to deal with them (Anderson 1999). There is also the problem of implementing policy given the presence of distortions in other markets. Correcting a distortion in one market may worsen a problem in another market. These external effects should be considered in assessing the effects of regulatory responses.

² A further issue is what is the appropriate level of government at which to apply the instrument. See Feketekuty (2000) for a discussion of the subsidiarity principle in this context.

doing nothing, options for dealing with problems of asymmetric information include:

- developing education campaigns;
- providing access to other sorts of legal remedies;
- introducing liability laws;
- promoting voluntary standards or codes of practice; and
- adopting technical regulations on service providers.

Similar lists can be devised for remedying other market failures. Coghlan (2000) presents a classification system that could be applied to all forms of regulation (see chapter 3 in this volume). His regulatory spectrum ranges from no action to ‘black letter law’ (box 3.3). Therefore, the choice of instrument requires an appropriate assessment of their respective costs and benefits.

The costs include those of:

- establishing and operating the system, including compliance costs incurred by government and firms having to comply with the regulation;
- any distortions, including side effects, such as on innovation or competition, that might be induced in markets where the regulation is applied;
- distortions in markets where tax revenue is raised to finance their development and operation; and
- the effects on existing distortions in other markets.

The benefits include the scale and frequency of market failures avoided. In other words, the correct choice of instrument requires a detailed specification of the objective. Problems that are likely to occur more often and with greater cost are likely to warrant different forms of intervention from problems having less significance. The instrument chosen should be proportional to the significance of the distortion.

Private interest forces and impacts

Services might also be regulated in the interests of well-organised groups who either initiate or capture the regulatory process. For example, pressure from private interests might:

- cause regulation to be introduced where none would otherwise be justified; or
- lead to the government imposing an inappropriate regulatory instrument according to the relevant tradeoffs.

It is important to note that regulation in the interests of these well-organised groups could be either more, or less, onerous than is socially optimal. Regulatory arrangements for dealing with negative externalities might be less stringent than optimal, while those dealing with asymmetric information might be tighter than optimal.

Inappropriate regulatory responses may also occur because of errors in decision-making processes. Pressures from various private interests can also influence the bureaucratic system that designs and implements the regulatory arrangements, including the attitudes to the cost of errors associated with regulation that is too light compared to that which is too tight. Coghlan (2000) reviews mechanisms designed to reduce the chance of error in regulatory design (see chapter 3 in this volume).

Further, given the choice of two regulatory structures that yield the same private benefits, governments might adopt the more obscure regulatory structure as a consequence of private interest pressure. The level of obscurity is already high for many of the policies adopted in response to market failure problems. More obscure instruments might also be more expensive to manage and maintain, which could involve an additional cost to private interests. However, more obscure instruments incur higher costs of action by countervailing interests, a feature that adds to their attractiveness to the interest groups favouring their application.

2.3 Costs of inefficient regulation

The assessment of the costs of inefficient regulation will depend on the circumstances of the case. There is a large literature concerning the choice of instrument for correcting externalities. There is also an emerging literature on the impact of inefficient choices of regulatory regimes to deal with competition policy issues. Warren and Findlay (2000) summarise one approach (see chapter 14 in this volume).

To illustrate the issues involved in assessing the costs of inefficient regulation, the remainder of this paper focuses on instruments addressing asymmetric information problems. There are interesting parallels between these issues and the growing literature on standards for goods in international trade (Maskus and Wilson 2000), although standards are only one of the menu of options noted earlier.³

³ There are also some interesting special issues in network industries. As Maskus and Wilson explain, standards are valuable in network industries to promote the growth of users of particular networks. There are significant network externalities, since the value of connection to any new user depends on the number of existing users. Standards for interoperability help deal with the

Researchers are developing techniques to assess the efficiency of regulatory instruments, including the level and degree at which the instruments are set (eg the level of standards). This paper focuses on evaluating one set of costs of inefficient regulation — the effect on trade and investment. The usual starting point for this type of assessment is the effect of a discriminatory regulatory system on the costs of a foreign supplier.

Higher marginal costs

A regulatory practice applied equally to both domestic and foreign firms may increase the marginal costs of both sets of firms. In addition, these costs are likely to be higher than necessary if the regulation is inefficient. Taking this as the starting point for regulatory reform, both consumers and producers will then gain from the introduction of more efficient regulatory structures. The distribution of benefits will depend on the elasticities of supply and demand.⁴ The important point is that the total welfare gain from removing the burden of inefficient regulation can be substantial. This is because, for example, a standard that has this impact is not generating rents or revenues to the government (in other words, the rectangle in the welfare analysis is not a transfer).

It is also useful to consider the impact of regulation within a dynamic setting. For example, a set of regulatory arrangements might be developed to solve a problem of asymmetric information. Over time, regulations can become more stringent and act as a barrier to entry as the regulatory instruments are captured by industry. So what was initially an appropriate policy response to a problem of asymmetric information may become a source of inefficiency. New technologies and other changes in the market might render some initial concerns about market failure irrelevant. Thus, regulations designed to suit an earlier period may later have perverse effects. These examples highlight the dynamic consequences of regulation. Therefore, regular reviews of regulatory policy are valuable to ensure that objectives remain valid and that the policies remain the most appropriate way of achieving those objectives.

risk of under-provision of such network services in a private market. On the other hand, private providers might also use standards to give themselves some strategic advantages. An assessment of these issues is reserved for another occasion.

⁴ The literature works over in detail issues relating to the impact of a shift in the supply curve brought about by technical change in agriculture. See Alston and Pardey (1996, chapter 5) for a survey.

Discriminatory effects

Of more interest here is the case where inefficient regulation impacts differently on the marginal costs of domestic and foreign suppliers. For example, the cost imposed on foreign suppliers might be higher because they have to satisfy an accreditation process in both home and export markets.

Consider the following assumptions:

- the foreign- and domestic-sourced services are perfect substitutes;
- the foreign supply curve is perfectly elastic (ie the services importer is a ‘small country’); and
- the domestic supply curve is upward-sloping (as in the standard model of the effects of a tariff).

Then, removing a discriminatory regulatory practice applying to foreign firms would lower their supply curve resulting in:

- a decrease in domestic prices;
- an increase in consumer surplus;
- a decrease in the surplus of domestic producers; and
- a large net increase in welfare, because the resource cost of meeting the standard by foreigners is now avoided.⁵

Maskus and Wilson (2000) illustrate the case of an upward-sloping import-supply schedule that shifts to the right as a result of removing the impediment on foreign supply and creates benefits for both foreign exporters and domestic consumers. At the same time, the domestic price falls (assuming the domestic supply curve is also upward-sloping) and local producers lose some producer surplus.⁶

Shifting demand curves

One complication is that this approach assumes that the consumer regards the foreign and local products as perfect substitutes. But demand for foreign-sourced

⁵ In the analysis of the effects of removing a tariff under similar assumptions, the loss of government revenue is a transfer to consumers within the economy applying the tariff. In the case of removing a discriminatory regulation, the saving in implementation costs (the equivalent of the tariff revenue in the tariff analysis) is a gain to the importing country.

⁶ Neven (2000) illustrates the case where the foreign supply curve is upward-sloping, but the domestic supply curve is perfectly elastic. The removal of the discriminatory regulation increases the market share of foreign suppliers without changing the domestic price.

services could increase as a result of meeting the domestic standard (Ganslandt and Markusen 2000). In other words, consumers actually regard domestic and foreign-sourced services as imperfect substitutes. The demand curve for foreign supplied services increases as a consequence of demonstrating a capacity to meet domestic standards.⁷

Imperfect competition also matters

Regulation will have different impacts in imperfectly competitive markets. Neven (2000) models the effect of discriminatory standards in a market with a specific number of domestic firms and one foreign firm. In the case of price competition, Neven finds that the imposition of a standard that raises marginal costs of the foreign firm (but not the domestic firms) increases the price of the output of the foreign firm and of the domestic firms, but increases the latter by less.⁸ The increase in price is larger:

- the smaller the number of firms — in other words, the higher the degree of domestic rivalry, the lower the protection induced by a discriminatory standard; and
- the larger the degree of substitution between products — that is, the lower the degree of differentiation.

Services, because of their nature and the manner in which they are produced and delivered, are likely to be relatively highly differentiated and, if so, the protective effect of a discriminatory standard is likely to be lower than otherwise.

Higher fixed costs

Perhaps an even more likely effect of regulation is that it will impose a one-off set-up cost on entrants. This cost might be higher for foreign than for local suppliers, for example, when foreign suppliers have to establish their credentials twice, once in the home market and again in the export market. Such higher fixed costs affect

⁷ A further effect of meeting domestic standards is that the degree of substitutability between domestic and foreign-sourced products might increase. The effects outlined in this paragraph can be taken up in more detailed modelling work.

⁸ Neven (2000) also finds that the value of imports could rise or fall with the imposition of a discriminatory standard. He comments that this uncertainty means that the value of imports is not a robust indicator of the extent of protection. The Neven model holds the number of firms fixed and so could be interpreted as applying in the short run. Baldwin (2000) stresses the impact of a discriminatory liberalisation of technical barriers to trade on the location of production. That is, liberalisation can influence the number of firms in the home and partner country between which a technical barrier is reduced compared to the number of firms in the rest of the world.

outcomes in imperfectly competitive markets — they reduce the incentives to enter, reduce the number of firms and raise the mark-ups on marginal cost.

On the other hand, the presence of domestic standards, which foreign firms are asked to meet, might also lower the costs of market entry for foreign firms. They reduce the costs to foreign firms of discovering information about the preferences in the importing market. As already noted, meeting them also provides an indicator of product quality. However, a non-discriminatory standard can make the same contribution at even lower cost.

2.4 Strategies to improve regulation

Measurement

A first step towards reducing the adverse impact of regulatory structures on trade and investment is to measure their effect. As is the case in other work on non-tariff barriers, measurement could focus on characterising regimes and then assessing their impact on prices, costs, profits and/or quantities.

Characterising the regime

The OECD has recently completed a sophisticated study characterising regulatory regimes and comparing them between members (Nicoletti, Scarpetta and Boylaud 2000). Their focus was the regulation of product markets in 21 OECD countries. However, the same approach can be used to develop indicators of regulatory arrangements for the service sector. The steps involved were:

- collecting basic data drawn from questionnaire responses and member economy publications (survey data accounted for 90 per cent of the information used);
- classifying the basic data according to three criteria:
 - economywide or industry-specific;
 - type (eg state control over business firms, barriers to entrepreneurial activity, barriers to international trade and investment); and
 - function (administrative procedures and other economic provisions);
- codifying the qualitative information or the subdivision of quantitative data into classes before normalising the scores on a common scale, say from zero to 6 (with the least restrictive economy in the sample given a score of zero and the most restrictive a score of 6), paying attention to the hierarchy between regulations; and

-
- applying statistical procedures to determine the best aggregates of the scores relating to individual regulatory instruments.

These steps are similar to those used in the studies of the service sector summarised by Warren and Findlay (2000) (see chapter 14 in this volume). However, the weights used in these latter studies to aggregate instruments were based on judgements of the researchers of the relative importance of different restrictions. The weights could also be derived from analysing the sequence of decisions that a firm might make in deciding whether or not to enter a particular market.

Nicoletti, Scarpetta and Boylaud (2000) and others developed a structure that incorporates all regulatory barriers to entry, not just those applying in a differential fashion to foreign firms. In terms of the structure of the GATS, it therefore isolates not only *national treatment* issues, but also appropriately covers *market access* issues.

What are the costs?

The simplest measure of the impact of a discriminatory regulatory regime is the extent to which it increases the costs of foreign firms. The extent to which these costs affect prices in the domestic market depends on the parameters highlighted by Neven (2000).

One approach to getting this information with which to examine the effects of regulatory structures is to survey firms (eg OECD 1999 summarised in Maskus and Wilson 2000). Maskus and Wilson (2000) highlight the value of further survey work in their specification of a research agenda on the impact of standards.

In subsequent work, and following the steps in studies reviewed by Warren and Findlay (see chapter 14 in this volume), these price impacts can be related to the policy measures developed in the manner outlined above. This step is a further test of the relevance of the aggregation procedures used to construct the summary indicators. However, one of the challenges is to adopt the appropriate definition of the market to that affected by the regulatory arrangement. The best market definition may not always be applicable because of a lack of appropriate data.

A further step relating policy measures to market outcomes is to derive various wedges that can be applied in general equilibrium modelling of the impacts of reform.⁹ Past studies, some of which are reviewed by Maskus and Wilson (2000),

⁹ Dee, Hanlson and Phamduc (2000) employ such techniques to model the effect of liberalising international trade in services using the FTAP model (see chapter 15 in this volume). The price

have chosen arbitrary reductions in border costs to capture the effects of reform. For example, trade costs are cut by 2.5 per cent or the elasticity of substitution between items traded by economies participating in the adoption of a market integrating measure is changed.

Linking regulation to trade outcomes

Another approach is to use econometric techniques to link the presence of standards or regulation to trade outcomes. Maskus and Wilson (2000) review recent studies, including one that uses a gravity model of bilateral trade flows. In that study, standards are divided into country-specific standards and bilaterally-shared standards. Standards were measured by counts of documents specifying the standard. The approaches to measuring restrictiveness in services trade and investment that the studies summarised in Warren and Findlay (2000) employed are likely to yield more interesting results than document counts in these sorts of studies (see chapter 14 in this volume).

Another issue in empirical work focusing on trade volumes or values is to establish expectations about the effects of removing discriminatory barriers on trade. Neven (2000) points to the problems of predicting the change in the value of imports when the impact of discrimination is modelled by a change in marginal costs. Baldwin (2000) also presents a case where market integration as a result of removing discriminatory standards or rules actually causes trade to cease.

However, the important point is that market integration in the model that Baldwin (2000) discusses also causes prices to fall. The difficulty of predicting quantity and value effects suggests that a focus on price impacts is more insightful.

Policy reform

Policy reforms worth considering may involve unilateral action or international cooperation. Coghlan (2000) demonstrates how modern processes of design and evaluation can lead to more efficient regulatory arrangements (see chapter 3 in this volume).

Role of international cooperation

Benchmarking domestic regulation against principles from international agreements can support the domestic reform process. International cooperation provides a way

wedges used were derived in Hanslow et al. (2000) using some of the studies summarised in Warren and Findlay (2000) (see chapter 14 in this volume).

of reducing the trade costs associated with regulation. For example, harmonisation of standards, in particular, the adoption of common international standards, certainly has that effect.¹⁰ But agreement on conformity also matters. For that reason, mutual recognition agreements may reduce trade costs.

Meeting international obligations is one of the characteristics of ‘good regulation’ stressed by Coghlan (box 3.6). International agreements — commitment to international standards and mutual recognition or the equivalent on conformity testing¹¹ — can also be used within the political processes to reduce the risk of regulatory structures being captured for private interest objectives. These agreements can be pursued at both a regional and global level. The next section examines the principles laid out in the GATS on issues of domestic regulation. An interesting subject for further research would involve assessing the treatment of domestic regulatory issues applying to services within regional trading arrangements — for example, North American Free Trade Association (NAFTA), Closer Economic Relations (CER) and Association of South East Asian Nations (ASEAN) Free Trade Area (AFTA) — and their relation to GATS commitments and the Asia Pacific Economic Cooperation (APEC) architecture.¹²

The contribution of the GATS

Article VI of the GATS requires members to develop disciplines on domestic regulation. The Article belongs to the section on general obligations and disciplines in the GATS. Briefly, the article covers measures relating to qualification requirements and procedures, technical standards and licensing requirements and procedures. It requires that such regulations should be:

- based on objective criteria;

¹⁰ Attention to the differences between voluntary and mandatory standards is important, since respondents in one survey reviewed by Maskus and Wilson (2000) indicated that meeting the terms of voluntary standards can be a larger issue for market access than meeting mandatory ones.

¹¹ There are also different approaches to conformity assessment. One approach assesses the capacity (eg equipment and personnel) to produce services before suppliers enter the market. The other approach is called a supplier declaration regime, where the supplier takes responsibility for ensuring that services made available meet the regulations. The latter has a number of advantages to suppliers. In all cases, some post-market surveillance may be valuable. Scoring or evaluating regulatory regimes should also take into account the conformity assessment process.

¹² In their declaration after the Darwin meeting in June 2000, the Ministers Responsible for Trade said that sub-regional trading arrangements should not only be consistent with WTO rules, but they should also be consistent with ‘APEC architecture’ and ‘supportive of APEC goals and principles’. According to the Osaka Action Agenda, the APEC principles relevant to regional arrangements include requirements on transparency and accession.

-
- transparent;
 - not more burdensome than necessary to ensure the quality of the service (the so-called necessity test); and
 - in the case of licensing, not in themselves a restriction on the supply of the service.

The WTO Secretariat prepared a note on the issues involved in the development of these disciplines, which remains work-in-progress for the Council for Trade in Services. So far, the work focuses on necessity and transparency issues. The Secretariat's interpretation is that there is nothing in Article VI that 'suggests its disciplines were to be limited to services on which specific commitments were undertaken' (ie they apply regardless of specific commitments).

The application of the necessity test is an important issue.¹³ The steps laid out by Coghlan (2000) for the development of good regulation are relevant (see chapter 3 in this volume). One GATS member could build on those procedures a strong rebuttal of a claim by another member that the necessity condition had not been met.¹⁴

Horizontal disciplines are a good idea

It would be worthwhile developing a horizontal approach to managing regulatory practices (ie an approach that covers many or all service sectors). Mattoo (2000) argues that a horizontal approach is useful because it:

- economises on negotiating effort;
- reduces the chances of capture by interests associated with any one sector; and
- achieves a broad coverage.

Mattoo (2000) also notes that a horizontal framework does not rule out sector-specific disciplines. Some principles are important and can be applied horizontally (eg those about the value of an independent regulator or the importance of not inhibiting competitive behaviour in the design of regulatory systems). Some other aspects of regulatory structures might have to be negotiated sector by sector, because of the nature of the production process or the technology involved.

¹³ Some of the discussion of the design of a necessity test stresses the criterion of minimal trade impact (WTO 1999a). However, more general principles for necessity will be derived from the assessment of the welfare effects of intervention usually observed in a second best situation.

¹⁴ Feketekuty (2000) makes a number of suggestions about how Article VI on domestic regulation might be strengthened.

Horizontal rules might also be used to decide what aspects of a regulatory system would have to be negotiated on a sectoral basis.

There may still be an issue about the efficiency of implementing regulatory practices. Mattoo (2000) stresses that, even if the multilateral disciplines on regulatory practices can be strengthened, the capacity in developing economies, in particular, to benchmark policy practices will still have to be enhanced. But even if the required architecture is in place, its management might appear to be relatively costly in developing economies (eg if the implementation process is relatively capital or skilled labour intensive). High costs arising from these circumstances would have to be distinguished from inefficient implementation of the administrative processes.

2.5 Conclusion

The risk of market failures in the service sector, such as those associated with asymmetric information, externalities or market power, leads to deliberation about the correct regulatory response. Efficient responses to the problems of market failure can be designed. The problem is that the management of such responses can be distorted or biased to different degrees by pressure from interest groups that inevitably gain or lose from intervention in the market place. A further important consequence of a biased choice of regulation is that it can create a degree of protection from foreign competition for local service providers.

Avoiding protectionist outcomes will lead to significant gains from regulatory reform. These gains include not only the gains to the importing economy as a whole, but also the gains to foreign service providers from the removal of discriminatory burdens associated with bad regulatory choices.

While it is possible to identify, at least in principle, the gains from reform, change is not automatic. A deeper assessment of the gains from reform, rather than just an in-principle argument, would be useful. A first step in the design and management of the reform process is to assess the regulatory arrangements and examine their impacts, including those on trade and investment flows.

The characterisation and assessment of regulatory structures is difficult, but not impossible. Highly sophisticated techniques for doing so are being developed. They are also being applied to particular markets and whole economies.

The next step is to define a set of principles to direct the reform program. The issue is not simply how to remove the regulatory arrangements. The problem is to find the best regulatory response to each market failure problem. The principles laid in the

GATS are a useful guide to direct domestic reform. International cooperation at regional, as well as global, level can also help maintain the momentum for change.

The broad outlines of a valuable set of principles to direct reform in a manner that achieves better and less trade distorting outcomes are available, but challenges remain in their implementation. As the Australian experience illustrates, the chance of making an effective response to these challenges is increasing, as economies try to develop more rigorous processes for regulatory review and re-design.

References

- Alston, J. and Pardey, P. 1996, *Making Science Pay*, AEI Press, Washington DC.
- Anderson, K. 1999, 'Agriculture, developing countries and the WTO Millennium Round', Discussion Paper no. 99.28, Centre for International Economic Studies, University of Adelaide.
- Baldwin, R.E. 2000, Regulatory protectionism, developing nations and a two-tier world trade system, Paper prepared for the Brookings Trade Forum, Washington DC, 27–28 April.
- Coghlan, P. 2000, 'The principles of good regulation', Chapter 3 in Productivity Commission and Australian National University, *Achieving Better Regulation of Services*, Conference Proceedings, AusInfo, Canberra.
- Dee, P., Hanslow, K. and Phamduc, T. 2000, 'Modelling the liberalisation of services', Chapter 15 in Productivity Commission and Australian National University, *Achieving Better Regulation of Services*, Conference Proceedings, AusInfo, Canberra.
- Feketekuty, G. 2000, 'Regulatory reform and trade liberalization in services', in Sauv e, P. and Stern, R. (eds), *GATS 2000: New Directions in Services Trade Liberalization*, Brookings Institution Press, Washington DC.
- Ganslandt, M. and Markusen, J.R. 2000, Standards and related regulations in international trade: a modelling approach, Paper prepared for the World Bank Workshop on 'Quantifying the trade effects of standards and technical barriers: is it possible?', Washington DC, 27 April.
- Hanslow, K., Phamduc, T., Verikios, G. and Welsh, A. 2000, Incorporating barriers to services trade into the FTAP database, Research Memorandum Cat. no. MC-59, Productivity Commission, Canberra.
- Maskus, K.E. and Wilson, J.S. 2000, Quantifying the impact of technical barriers to trade: a review of past attempts and the new policy context, Paper prepared for

-
- the World Bank Workshop on ‘Quantifying the trade effects of standards and technical barriers: is it possible?’, Washington DC, 27 April.
- Mattoo, A. 2000, ‘Developing countries in the new round of GATS negotiations: towards a pro-active role’, *World Economy*, vol. 23, no. 4, pp. 471–91.
- Neven, D.J. 2000, Evaluating the effects of non-tariff barriers: the economic analysis of protection in WTO disputes, Paper prepared for the World Bank Workshop on ‘Quantifying the trade effects of standards and technical barriers: is it possible?’, Washington DC, 27 April.
- Nicoletti, G., Scarpetta, S. and Boylaud, O. 2000, ‘Summary indicators of product market regulation with an extension to employment protection legislation’, Economics Department Working Papers no. 226, OECD, Paris.
- OECD (Organisation for Economic Co-operation and Development) 1999, ‘An assessment of the costs for international trade in meeting regulatory requirements’ TD/TC/WP(99)8/FINAL, OECD, Paris, March.
- Vousden, N. 1990, *The Economics of Trade Protection*, Cambridge University Press, Cambridge.
- Warren, T. and Findlay, C. 2000, ‘Sins of commission and omission: measuring regulatory impediments to trade in services’, Chapter 14 in Productivity Commission and Australian National University, *Achieving Better Regulation of Services*, Conference Proceedings, AusInfo, Canberra.
- WTO (World Trade Organization) 1999a, ‘Article VI:4 of the GATS: disciplines on domestic regulation applicable to all services’, S/C/W/96, Council for Trade in Services, WTO, Geneva, 1 March.
- 1999b, ‘Report of the working party on domestic regulation to the Council for Trade in Services’, S/WPDR/1, Working Party on Domestic Regulation, WTO, Geneva, 20 October.

Discussant — *Patrick Messerlin*

Institut d'Etudes Politiques de Paris

The appealing point in Christopher Findlay's paper is the focus on domestic regulatory reform. For some time to come, most regulatory reform will come from within individual countries, rather than through international negotiations such as the GATS or General Agreement on Tariffs and Trade (GATT).

These comments address the literature behind the paper, before adding a few points based on the European Community's experience in liberalising services.

D2.1 Economic literature behind the paper

The paper correctly focuses on domestic regulation, although an alternative focus may be slightly more appropriate. The economic justification for the approach taken here goes back to what is known in trade theory as distortion theory. Distortion theory emphasises two points that are relevant. First, Christopher Findlay mentions the cost of regulation, but Bhagwati (1971) also mentions the opportunity cost of regulation. In a two-good and two-factor model, a regulation affecting good A has a cost reflected in good B. Although Christopher Findlay's paper briefly mentions the second point, the principle of compensation — that the policy instrument should compensate for the distortion — it is crucial to whether policy instruments increase overall social welfare. The existence of a distortion is not enough to justify intervention. A proportionality test would also be extremely helpful. This would assess whether the response being advocated was proportional to the problem being solved.

There are two reasons why distortion theory has, in a sense, been lost.

First, Bhagwati ignores the political economy views that this paper mentions. The theory is naively presented. Distortion theory suggests that a production subsidy is better than a tariff, but what if the production subsidy is not transparent? An opaque instrument is easier to capture than a transparent one. Opaque instruments can also lead to over-compensation, resulting in a welfare loss. Therefore, the theoretically 'better' instrument may not be the best instrument to use in practice.

The theory of Directly Unproductive Profit-Seeking (DUP) extends distortion theory to include some political economy points of view. However, not all distortions result in transfers of rents or taxes from one member of the economy to another. Some distortions do not involve transfers between agents. With the Common Agricultural Policy (CAP), many small French farmers receive a negligible subsidy, but face a large administrative burden under the scheme. For

such farmers, the CAP is a pure waste. Thus it is extremely important to apply a proportionality test to proposed solutions in the presence of distortions.

Second, the dynamic aspects of regulation should also be considered. The paper focused on the impact on foreigners, but there are also impacts of regulation on domestic producers in a dynamic setting. An example is the French retail industry. In the beginning, rules about entering the retail market (on size, location etc) increased the cost to new entrants, but when retailers finally got into the system, they used these rules for their own sake, consequently increasing domestic costs over time. Regulations often introduce imperfect competition and raise costs for incumbents in expanding their business.

D2.2 Policy-orientated aspects

The experience of the European Community in the liberalisation of services in general, and telecommunications in particular, is that success has generally come from technical progress, not because of what the Community or Member states have done. Technical progress is pro-competitive in some cases and anticompetitive in others. It has been good for competition in banking and capital markets, but not so for competition among airlines — the development of computerised airline reservation systems has allowed airlines to manage the so-called liberalisation process and to collude at the expense of consumers (see chapter 6 in this conference volume). It is important to understand the link between technological change and regulation and/or liberalisation. In the 1880s, technical progress in shipping (the shift from sail power to the steam engine) led to lower freight costs and, with it, increased US competition in European agricultural markets. If the ongoing liberalisation resulting from technical progress is not recognised, governments may increase tariffs, as in last decade of the 19th century, in order to deal with the problem.

Christopher Findlay mentions that networks are important and that they are at the core of things to be regulated. However, in some cases, networks are exploding. The TV industry may not be recognisable ten years from now, because of the internet. Clearly, this has implications for existing policies and instruments (eg subsidies and quotas). A contrary example is nuclear electricity. In France, nuclear energy accounts for 80 per cent of electricity generation. It cannot be privatised or sold, as nobody would want to buy it. The problem is how to introduce competitive supply into such a situation. It will take at least 10 to 15 years before competition in the electricity network will come to fruition.

Policy reviews are the key. The right instrument implemented at the wrong time can have perverse consequences. This is why policy reviews are so important, because

they decide the timing of the decision. European competition policy, for example, has been very active in the field of audiovisual services, very active and very counterproductive (see chapter 12 in this volume). Good instruments applied at the wrong time can be just as perverse. This is true of the European ‘anti-subsidy’ policy that favours large firms. The policy actually has the perverse effect of being more discriminating than the presence of a subsidy because of its high administrative burden.

Christopher Findlay’s paper favours mutual recognition agreements and a horizontal approach to liberalising trade in services. However, mutual recognition agreements and the horizontal approach are contradictory and this is a major problem. The EC experience of mutual recognition agreements is that they involve negotiation about rules and provisions. So, by definition, governments negotiate mutual recognition agreements on a sectoral basis and not on a horizontal basis. Thus, there is a driving force in mutual recognition agreements towards sectoral agreements, and against horizontal or Community-wide agreements. Sectoral agreements do not provide for the tradeoffs between sectors that a horizontal approach does.

Reference

Bhagwati, J.N. 1971, ‘The generalized theory of distortions and welfare’, in Bhagwati, J.N. et al. (eds), *Trade, Balance of Payments and Growth: Essays in Honor of Charles P. Kindleberger*, North-Holland, Amsterdam.

General discussion

The discussion focused on:

- rationales for regulating services;
- a horizontal versus sectoral approach to regulatory reform; and
- regional versus multilateral disciplines.

Rationales for regulating services

One participant suggested that services were different from goods and, as such, warranted different regulatory approaches. The participant contended that it was often a lot more difficult to define the issues for services and that services were less likely to be traded internationally than goods — countries impose substantially different requirements for many service industries (eg health, education, tourism, accounting and the law). The participant also commented that national sovereignty was an important issue that needed to be considered. Countries may want to differentiate their domestic markets and the products in their markets, which may have ramifications for the regulatory regime.

Another participant suggested that the three rationales presented in Christopher Findlay's paper for government intervention — market power, externalities and information failure — could give excuses to intervene in all circumstances. There were dangers in focusing on information requirements, and remedying them could distort trade. Informational requirements would generally solve themselves, as people found ways of attracting consumers and emphasised the positive aspects of their services.

Christopher Findlay responded that there was still a need to consider information asymmetry and market failure as rationales for regulation. The business community thought that these issues were really important, especially in relation to standards and management of these information problems. The key was to address these problems in ways that did not promote their exploitation.

A horizontal versus sectoral approach to regulatory reform

In response to a question about his paper, Christopher Findlay stated that the aim of a horizontal approach was to develop principles to apply at the broad sectoral level. Most of the work on mutual recognition related to conformance rather than to the standards themselves, and, in the area of conformance, it was unclear whether there was a conflict between mutual recognition and the horizontal approach. However,

Christopher Findlay agreed that there was a need to maintain a discipline or constraint on sectoral negotiations to avoid many of the pitfalls associated with mutual recognition agreements.

Regional versus multilateral disciplines

Christopher Findlay was asked whether there was much evidence that regionalism had generated outcomes inimical to third countries. In the NAFTA negotiations, such as those on professional services, for example, the negotiators understood the difficulty in making progress within a multilateral setting, but nevertheless applied GATT disciplines on goods to services in a regional setting. The limited progress made within NAFTA on professional services was still positive. Furthermore, regional rules typically followed multilateral practice, so the risk was not as great as expected.

The same participant was more concerned that regionalism was only weakly disciplined by the WTO. For example, Article V (Economic Integration) of the GATS was an extremely weak discipline and therefore regional agreements would not necessarily be informed by the multilateral disciplines developed under Article VI.4 (domestic regulation). In addition, the disciplines on domestic regulation would not apply strongly to regionalism. There was a need to be careful about promoting outcomes on a regional basis.

Christopher Findlay noted that NAFTA was a good example where regional agreements could achieve progress that may be difficult under the multilateral system. The literature on regional agreements did not deny that view. But the problem with regional arrangements was that they ran the risk of stalling. The political economy for pursuing further reform was easily lost when everyone was cosy within a regional arrangement.

There seemed to be evidence that more countries were interested in regional agreements and these agreements covered not only tariffs, but also business and trade facilitation matters. But there definitely needed to be some sort of discipline on these agreements. The multilateral system as it currently works was not going to provide this. The GATS applied very weak disciplines to that regional process. However, some higher level discipline was needed to stop the regional approach losing momentum.

3 The principles of good regulation

Paul Coghlan[†]
Productivity Commission

3.1 Introduction

There are two certainties in life: death and regulation. Regulation is integral to the equitable and effective operation of society, and its effects are so pervasive, that it is appropriate to ponder the general principles to which regulation should adhere.

A sensible start might be to ask the legal professionals and legislative experts what these general principles are — after all, they have been in the regulatory ‘business’ for centuries. From their perspective, examples of good regulatory principles include:

- clarity and precision of legislation;
- adherence to the general objects and spirit of the empowering legislation;
- not unduly trespassing on personal rights and liberties;
- not reversing the onus of proof in criminal proceedings; and
- ensuring protection from self-incrimination (Bridgman and Davis 2000, p. 62; Douglas and Jones 1999, pp. 308–15).

While these are all laudable principles, they do not cast much light on the central concerns of economists and policy analysts — whether regulation is an effective and efficient means of achieving policy objectives, and whether it provides a net benefit to the community. This paper develops a more satisfactory set of principles of good regulation by:

- setting down different types of ‘regulation’ (section 3.2);
- considering if, and when, governments should regulate (section 3.3);

[†] The views expressed in this paper are those of the author and do not necessarily reflect those of the Productivity Commission.

-
- deciding what forms of government action might best be adopted (section 3.4); and
 - examining features that characterise poor or ineffective regulation (section 3.5).

This points to some essential characteristics and principles of good regulation (section 3.6) that should provide useful benchmarks against which to assess current practices.

Sections 3.7 and 3.8 describe a process endorsed by the Commonwealth Government of Australia and implemented by the Office of Regulation Review to help ensure that the design of Commonwealth regulations reflects such principles.

3.2 Types of regulation

There are four broad types or ‘tiers’ of regulation (box 3.1).

Primary legislation comprises Acts of parliaments. While they can contain highly specific provisions, they are often expressed in quite general terms, setting down overall guidelines as to what the legislation is intended to do and delegating the implementation detail to the second tier — delegated legislation.

Box 3.1 Regulatory tiers

1. Primary legislation
 - Acts of parliaments
2. Delegated legislation
 - statutory rules
 - disallowable instruments
 - non-disallowable instruments
3. Quasi-regulation
 - codes of practice where there is a government role
 - standards, such as some Australian standards
 - guidelines
4. Administration of regulations

As indicated in box 3.1, delegated legislation can take many forms. The first two, statutory rules and disallowable instruments, are subject to limited parliamentary scrutiny, but considerably less so than for primary legislation. The other element of delegated legislation, non-disallowable instruments, is the responsibility of Ministers, relevant boards and officials.

The third tier is referred to in Australia as quasi-regulation. It covers those rules or arrangements where governments influence people to comply, but which do not form part of explicit government regulation (IDC on Quasi-regulation 1997, p. 7). (For example, there is an Electronic Funds Transfer Code of Conduct covering financial transactions effected with a card and PIN number; the Australian Securities and Investment Commission monitors the operation of this code.) Note, that each of the three elements of quasi-regulation listed in box 3.1 may be embedded in primary or delegated legislation, thereby transforming them into explicit government regulation.

The fourth tier is about how regulatory officials actually implement the regulation. Surveys of businesses often point to general acceptance of regulations, but highlight concerns about the difficulties with the way they are administered and with the associated compliance costs (Small Business Deregulation Task Force 1996b, p. 13).

These tiers of regulation highlight three points pertinent to achieving good quality regulation:

- the lower one goes down through the regulatory tiers, the fewer parliamentary and other quality control mechanisms there are;
- if more quality control is applied at the higher tiers, there will be pressure to shift more and more of the regulatory function to lower tiers; and
- the magnitude of economic impact does not necessarily diminish at lower tiers; indeed, administrative or quasi-regulatory arrangements may have substantially more economic impact than primary legislation.

From a global perspective, international treaties and agreements are increasingly influencing domestic regulation. The rapid growth of international trade in services and the global aspects of some environmental issues have contributed to this trend. It may be impossible to design domestic regulation in accordance with good principles, if its purpose is to implement an international treaty which fails to meet these principles. Thus, the application of good regulatory principles is a worldwide, not just a domestic, imperative.

3.3 To regulate or not?

A feature of democratic systems is that, when problems arise, voters say, ‘Why doesn’t the government do something about ... ?’ Politicians want to demonstrate that they have done something. Policy advisers and legislative drafters traditionally have served up regulatory solutions with minimal consideration of other options.

While challenging the traditional presumption favouring government regulation, an economist recognises that government intervention may be appropriate under certain conditions — when markets fail to produce optimal economic outcomes and for addressing environmental, equity and other social objectives that are unlikely to be met by market forces. But apart from these conditions, competitive market forces should, in general, produce optimal results. Kahn (1989, p. 18) put it succinctly:

Competition will weed out the inefficient and concentrate production in the efficient: it will determine, by the objective test of market survival, who should be permitted to produce; it will force producers to be progressive and to offer customers the services they want and for which they are willing to pay; it will assure the allocation of labor and other inputs into the lines of production in which they will make the maximum contribution to total output.

This approach received substantial endorsement in 1995, when all Australian governments signed the Competition Principles Agreement (CPA) (NCC 1998). One element of the CPA required all governments to systematically review their stock of legislation, focusing on elements that restrict competition. Such elements are among the most interventionist forms of government regulation — for example, bans on business ownership, barriers to starting new businesses or providing professional services, and prescribed quality and standards (NCC 1997, p. 4). The CPA requires that governments abolish such legislated restrictions on competition, unless it can be demonstrated that:

- (a) they provide a net public benefit; and
- (b) the government objectives can be achieved only by restricting competition.

The CPA specifies that any assessment of net public benefit (see (a) above) must take into account, among other things:

- government legislation and policies regarding ecologically sustainable development;
- social welfare and equity considerations; and
- the interests of consumers generally or of a class of consumers (NCC 1998 – CPA Clause 3).

The CPA set new ground rules for regulatory design, for two reasons. First, it reverses the onus of proof. No longer do those advising against regulation restricting competition have to overturn the status quo and prove that the removal of such regulation will result in a net benefit to the community. Rather, the onus now rests with those proposing the maintenance of such restrictions to *demonstrate* that they provide a net benefit to the *community*. Second, it must also be demonstrated that the government's specified economic and social objectives cannot be met by any means other than by restricting competition.

These principles of good regulation are highly pertinent to the regulation of services. Box 3.2 lists some examples of services where the Commonwealth Government is actively considering a range of regulatory options.

Each of those examples involves different aspects of competition — whether and where pharmacists can start a business, what type of digital TV must be broadcast and when, the quality of service that should be provided by a mediator, and so on. They all limit the way service providers can compete.

These examples highlight why good regulation of services is particularly important:

- services are the fastest growing sector of the economy and any unduly restrictive regulatory controls will sap prospects for employment growth and valuable productivity gains;
- services are often closely linked to rapid technological development which could be hampered by applying inflexible regulations;
- some services are highly mobile in a global sense and can move across international boundaries towards countries with the least onerous regulatory regimes; and

Box 3.2 Current examples where Commonwealth regulation of services is proposed or under consideration

1. Maintenance of controls over the location of pharmacists and (by implication) the services they offer.
2. Controls over digital television and datacasting services.
3. Regulation to protect the personal information held by businesses, including that collected from electronic commerce transactions.
4. Quality standards for persons involved in alternative dispute resolution (eg mediators).
5. Implementation of an access regime for postal services.

-
- the regulation of services may be inherently intrusive because it tends to focus not only on the end product, but also on the manner in which it is delivered.

3.4 What forms should government action take?

There is a variety of factors relevant to deciding what form of action government should take to address a problem:

- the extent of risk;
- the severity of the problem;
- the nature of the industry concerned; and
- the need for flexibility or certainty in any regulatory arrangement.

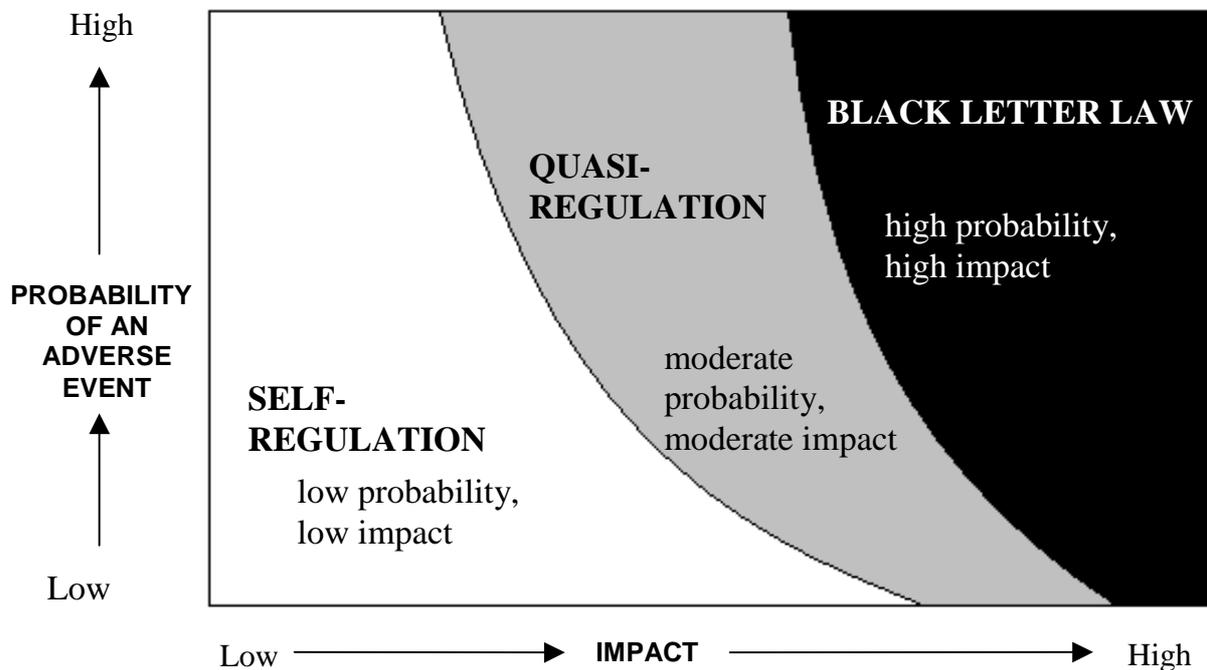
Relatively ‘light-handed’ regulation, such as industry agreed codes of practice, are likely to have fairly low economic costs, as they do not change the fundamentals of how an industry operates. In contrast, ‘heavy-handed’ regulation, such as restricting entry to markets, may substantially alter the structure and operation of an industry, and is likely to incur relatively large economic costs. It is an important principle, therefore, that regulation should be the minimum necessary to achieve the stated objectives. In general:

- matters where the potential damage is minimal might best be left to the market or industry itself to correct;
- matters with moderate probability and/or moderate impact may best be handled by a relatively ‘light-handed’ regulatory approach; and
- matters with high probability and/or high impacts may warrant highly interventionist regulation.

Figure 3.1 illustrates tradeoffs between risks and impacts and what form of government action might be appropriate.

For example, the probability of a mid-air collision between two aircraft might be very low, but, should it eventuate, the consequences are likely to be great. Such a matter is likely to justify explicit government regulation. In contrast, the risk of someone being overcharged for a supermarket item may have a moderate probability, but the impact is quite low — the diagram suggests such an event should be left to the industry and not be addressed by government regulation.

Figure 3.1 Notional relationships between risk and regulatory forms



As is particularly relevant to the provision of many services, recent rapid developments in electronic commerce and related technological advances add another dimension to the relationships illustrated above. Government regulation is inherently slow to be developed and subsequently to be changed, and typically is inflexible. As a general rule, therefore, explicit government regulation is likely to be less effective in sectors experiencing rapid rates of technological development or other changes. Paradoxically, the uncertainty associated with such rapid change may result in strident demands for more regulation. For example, rapid growth in many different forms of electronic transactions has raised demands for governments to regulate for protection of the privacy of information on individuals.

In practice, the relationships illustrated in figure 3.1 do not provide much help in resolving the question as to what would be the most appropriate form of government action when a problem arises. Using the privacy of personal information example, where such issues might lie along the probability/risk/impact axes depends very much upon the perceptions of individuals. Further, the risk of having one's spending patterns disclosed might be perceived quite differently from having one's health details disclosed.

With those qualifications in mind, it remains a sound general principle of good regulation-making to start at the 'no action' end of the regulatory spectrum when considering the most appropriate government response.

The regulatory spectrum

Box 3.3 sets out nine different forms that a government response to a problem might take, ranging from no action all the way to explicit government regulation. There are many potential intermediate steps. The first three are non-regulatory. The fourth is industry self-regulation. Items five and six cover industry codes of practice or standards and are typical examples of quasi-regulation. Moving further along the regulatory spectrum are government guidelines (item seven). Finally, items eight and nine are explicit government regulation.

One example of quasi-regulation is the supermarket scanning code, an industry-based code with government endorsement (type 5 in box 3.3). The code informs consumers about the price scanning technology used in supermarkets and provides assurances of its integrity. There are informal, accessible ways of dealing with customer complaints — such as providing an item free of charge if it is scanned at a price higher than the shelf price. The (then) Trade Practices Commission allowed its logo and endorsement to be used for the code.

Moving along the regulatory spectrum, the Australian Competition and Consumer Commission guidelines on pricing and the introduction of the goods and services tax (ACCC 1999) is an example of type 7 in box 3.3. Section 75U of the *Trade Practices Act 1974* (Cwlth) (TPA) prohibits price exploitation in relation to the ‘New Tax System’ changes. The ACCC’s guidelines provide general principles, information and practical guidelines on when pricing would contravene that part of the Act

Box 3.3 The regulatory spectrum

1. No government action or regulation.
2. Government action of a non-regulatory nature (eg information programs).
3. Agreements negotiated between industry and government.
4. Industry self-regulation.
5. Industry based code of conduct or standard with government endorsement.
6. Direct government involvement in the development and subsequent monitoring of an industry code of practice or standard.
7. Government guidelines to assist business meet legislative requirements by suggesting actions not specified in law.
8. Codes of practice and standards embedded in regulation.
9. ‘Black letter law’ — explicit government regulation.

A third example, in the dark end of the spectrum, is the Franchising Code of Practice, which is prescribed as mandatory under the TPA (see type 8 in box 3.3). Following well publicised concerns about small business persons losing life savings in ‘shonky’ franchise arrangements and earlier failure of a voluntary code of practice, this code was developed in a collaborative way between industry and government. This approach ensures a large degree of industry ‘ownership’ of the code with an assurance of coverage and enforceability provided by TPA backing.

Choosing from the regulatory spectrum

Of course, in practice, there are further distinctions that the rather simplistic picture in box 3.3 cannot capture, but that must be taken into consideration when choosing the most appropriate way to address any problem.

From an economic perspective, some forms of self-regulation may be an inferior option to explicit government regulation. For example, if a group of service providers, such as migration agents or medical specialists, were to self-regulate in order to provide clients with assurance of a good quality service, there may be a temptation for them to set standards which effectively exclude potential new entrants to the industry. Another more general example is that small businesses typically prefer the certainty of explicit government regulation, rather than the higher transaction costs often associated with other forms of regulation.

Further, the last category — ‘black letter law’ — has a wide range of different regulatory forms, some of which are much more interventionist than others, and thus more likely to be relatively costly. At the ‘lighter’ end of this group may be taxes, fees and fines, business registrations etc. At the ‘heavy-handed’ end would be prohibitions on any particular business activity, barriers to entry into certain industries, price controls and other operational restrictions on business, all of which constitute restrictions on competition.

A related consideration, which applies right across the regulatory spectrum, is the distinction between prescriptive (‘command and control’) regulation and performance/outcome-oriented regulation. Performance-based regulation is widely seen as appropriate for building flexibility into regulation and for encouraging innovation. By focusing on the outcomes required, rather than prescribing the precise means of achieving those outcomes, this approach gives business the opportunity to achieve regulatory objectives in ways that suit their needs and minimise compliance costs. For example, to protect workers from noxious fumes, prescriptive regulations may specify the type of ventilation system to be installed. Rather, if only the level of acceptable exposure were set, businesses might find that the best approach to reduce noxious fumes is to alter the mix of inputs and the

production processes, possibly bettering the acceptable level, at a lower cost than installing a ventilation system.

Clearly, the process of choosing the best form of action or regulation to address a particular problem is not straightforward. The following general guidelines are provided to officials developing or amending Commonwealth regulations (ORR 1998, pp. D4–D5).

Self regulation should be considered where:

- there is no strong public interest concern, in particular, no major public health and safety concern;
- the problem is a low risk event of low impact/significance; and
- the market can fix the problem itself — for example, there may be an incentive for individuals and groups to develop and comply with self-regulatory arrangements (eg for industry survival or market advantage).

The likelihood of self-regulatory industry schemes being successful is increased if:

- there is adequate coverage of the industry concerned;
- there is a viable industry association;
- there is a cohesive industry with like minded/motivated participants committed to achieving the goals; and
- there is evidence that voluntary participation can work — effective sanctions and incentives can be applied with limited scope for the benefits to be shared by non-participants.

Quasi-regulation should be considered where:

- there is a public interest in some government involvement in regulatory arrangements and the issue is unlikely to be addressed by self-regulation;
- government is not convinced of the need to develop or mandate a code for the whole industry;
- there are cost advantages from flexible, tailor-made solutions and less formal mechanisms, such as access to speedy, low cost complaints handling and redress mechanisms; and
- there are advantages in the government engaging in a collaborative approach with industry and industry having substantial ownership of the scheme.

Explicit government regulation should be considered where:

- the problem is a high risk event of high impact/significance (eg a major public health and safety issue);
- the government requires the certainty provided by legal sanctions;
- universal application is required;
- there is a systemic compliance problem with a history of intractable disputes and repeated or flagrant breaches of fair trading principles and no possibility of effective sanctions being applied; and
- existing industry bodies lack adequate coverage of industry participants, are inadequately resourced or do not have a strong regulatory commitment.

These guidelines are complementary to a formal testing of the effectiveness and the likely costs and benefits of different regulatory options discussed in section 3.7.

3.5 Features of poor regulation

Consideration of some of the characteristics of poor and ineffective regulation can provide further clues as to essential principles of good regulation.

High compliance costs

Even though businesses or individuals may have no disputes about the nature of particular regulations, there may be substantial concerns about unreasonably high costs incurred in complying with those regulations. Taxation law and associated regulations in Australia have been at the forefront of these concerns. A 1996 survey undertaken for the Small Business Deregulation Task Force (1996a, p. 3) revealed that, of the additional compliance costs incurred by small business (additional to commercial accounting processes), three-quarters were attributable to complying with taxation requirements.

Taxation regulation amply demonstrates a range of problems affecting many other regulatory regimes and which add to compliance costs:

- successive new ‘waves’ of regulation are required to patch up loopholes in existing regulation so as to proscribe tax avoidance;
- constant change in some aspects of tax regulation makes it particularly difficult to know what must be complied with and how to comply; and
- the regulation is complex and/or poorly expressed so that professional help must be obtained to help ensure compliance.

As a general rule, such areas of regulation are evident by the strong growth of ‘information asymmetry’ industries, such as taxation consultants, financial advisers and other experts.

Inaccessibility of the law

Regulation will not be effective, and certainly not efficient, if those who must comply have substantial difficulty in gaining knowledge of what it is they must comply with, and how they must comply.

Referring back to box 3.1 that set out four tiers of regulation, the second tier was delegated legislation. Some of that is tabled in Parliament, but not all of it. For the Commonwealth, there is no central register of all delegated legislation to facilitate access to the law.

The third category, quasi-regulation, is not readily accessible. Indeed, in cases where Australian Standards are embedded in regulation, it may be necessary to purchase the relevant standard (Standards Australia is not a government body, rather a private sector standards making organisation) in order to determine the detail of the regulation to be complied with.

In recent years, Australian governments have taken substantial steps to improve access to the law. A Business Entry Point (BEP) project has been developed to provide a wide array of information via electronic means to businesses in terms of licence and operating requirements and what regulations apply (DEWRSB 2000, p. 18). Nevertheless, the principle of accessibility needs to be given due weight early in the process of designing new regulation, so that it does not later become a problem.

Too much scope for discretion

It is always a difficult judgement as to how much detail should be embedded in regulation. If there is too little detail, those officials who administer the regulation may be given too much scope for discretion in interpreting the regulation. This effectively shifts power from the legislature to officials. There have been occasions when one office of a regulatory agency has given quite a different interpretation from an office in a different State of the same agency. In general, too much scope for interpretation of regulation by officials increases uncertainty for those regulated and raises compliance costs.

In contrast, very detailed and highly prescriptive regulation runs the risk of being inflexible by specifying how those regulated must go about complying. It would be

better to indicate what regulatory outcome must be met and then leave it to those who must comply to select the least-cost means of achieving that outcome.

'Thirty different governments'

This is the title of a 1994 report prepared about regulation in NSW (Sturgess 1994). In essence, those being regulated perceived a lack of consistency and coordination between different NSW government agencies. In attempting to comply with the law, businesses felt that they were confronted by many separate governments. Such perceptions have led some States to set up 'one-stop-shop' facilities to assist businesses.

Nationally, there are often perceptions of a lack of regulatory coordination between Commonwealth, State and Territory, and local governments. A good example is environmental regulation when domestic law implemented by the States sometimes is necessary as a consequence of an international environmental treaty entered into by the Commonwealth — such as greenhouse gas commitments.

Collateral damage

Some regulation, although well focused on the target being addressed, can cause widespread damage elsewhere. For example, if large revenue losses through the 'negative gearing' of residential investment prompted a removal of such tax provisions, there could be a substantial shift of investment to other forms, a shortage of rental properties and markedly higher rents. There could also be a distributional impact disadvantaging low-income households, which are more likely to rent than to buy a house. Full consideration of the impacts of regulatory proposals should help guard against such 'collateral damage'.

Failure to enforce the regulation

Clearly, no regulation will be effective unless it is enforceable, and ideal regulation will have the widespread support of those having to comply with it and so can be largely self-enforcing. In cases where more direct government pressure is needed to enforce the regulation with suitable penalties and other sanctions, adequate resources must be allocated to that task. If not, the regulation will be ignored and be ineffective, and that could contribute to a general culture of non-compliance with other laws.

3.6 Characteristics of ‘good’ regulation

The foregoing discussion of when governments should regulate, the different forms that regulation can take, and some features of poor regulations, can now be drawn together into essential characteristics of effective and efficient regulation.

1. Regulation must yield a *net benefit to the community*, not just to a particular group or sector.
2. Regulation must be *set to the minimum level necessary* to achieve the objectives and to avoid unnecessary restrictions. It should be targeted at the problem.
3. Regulation should be *integrated and consistent with other laws*, agreements and international obligations. Any restrictions on competition should only be retained if they provide a net benefit to the community and if government objectives cannot be achieved by other means.
4. Regulation should *not be unduly prescriptive* and, preferably, be specified in terms of performance or outcomes. It should be flexible enough to allow businesses some freedom to find the best way for them to comply and so it can adapt to changed circumstances.
5. Regulation should be *accessible, transparent and accountable*. Not only should the public be able to readily find out what regulations they must comply with, but the regulations must also be reasonably easy to understand and they should be fairly and consistently administered and enforced.
6. Regulation must be *clear and concise and communicated effectively*.
7. Regulation should be *mindful of the compliance burden* imposed, proportionate to the problem being addressed and set at a level that minimises compliance costs while still achieving the set objective.
8. Regulation must be *enforceable* and embody the minimum incentives needed for reasonable compliance. Adequate resources must be provided for monitoring and to ensure reasonable compliance.

These principles draw on those enunciated for all Australian governments by the Council of Australian Governments (COAG 1997) and those endorsed by the Commonwealth Government in *A Guide to Regulation* (ORR 1998). Wide ranges of other sources identify similar principles of good regulatory design, including the Victorian Office of Regulation Reform (Government of Victoria, undated), the Government of Canada (2000), and the OECD (1997).

3.7 Processes to promote good regulation

Governments in Australia and the OECD have become increasingly aware that improving regulatory regimes is an essential part of microeconomic reform. Regulatory reform has the potential not only to boost domestic productivity, but also to improve the ability to attract increasingly mobile economic activity. Many service industries are globally mobile. Over the past ten years, and particularly in recent years, many governments have implemented processes to promote good regulatory principles — most OECD member countries now have in place some form of regulatory quality assurance process (OECD 1997).

Having a sound process for making regulation is an essential prerequisite for obtaining good regulatory outcomes.

The Commonwealth Government has imposed quality control processes to ensure that regulation should only proceed if it results in net benefits to the community. Commonwealth departments and regulatory agencies are required to justify the need for explicit government regulation and to consider alternative ways of attaining policy objectives. A major vehicle for ensuring quality control is regulation impact analysis, which calls for an economywide perspective in identifying who benefits from the regulations, who incurs the costs and whether the regulation achieves its objectives without excessively burdening the community.

The Commonwealth Government has declared that a regulation impact statement (RIS) *must* be prepared and presented at two stages:

- Stage 1:* when the proposal goes to the decision maker, be that Cabinet, the Prime Minister, other Ministers or a decision-making board; and
- Stage 2:* when the proposal is tabled in Parliament (thereby making the RIS publicly available).

The elements of a RIS are shown in box 3.4. The RIS requirement forces policymakers to work through a sequential process of articulating the problem, assessing a range of options and recommending the best option or explaining why some other option is recommended. (The third element, setting out some viable options as to how desired objectives might be achieved, prompts officials to work along the regulatory spectrum outlined earlier in this paper.) Taken together, all the elements of a RIS essentially constitute a systematic policy development process designed to implement principles of good regulation.

The two stages at which RISs must be prepared are marked towards the bottom of figure 3.2, which depicts an ideal policy development process. However, the RIS process has the potential to make greater contributions to sound regulatory

outcomes, if also undertaken earlier in the policy-making process. The RIS prepared early can serve as a consultation document and draw out views and data from those likely to be affected.

The Government has made the Office of Regulation Review, which is part of the Productivity Commission, responsible for ensuring that all Commonwealth departments and regulatory agencies prepare and submit RISs of a satisfactory standard.

Box 3.5 provides a specific example of a RIS recently tabled in the Commonwealth Parliament concerning implementation options for an access regime for postal services.

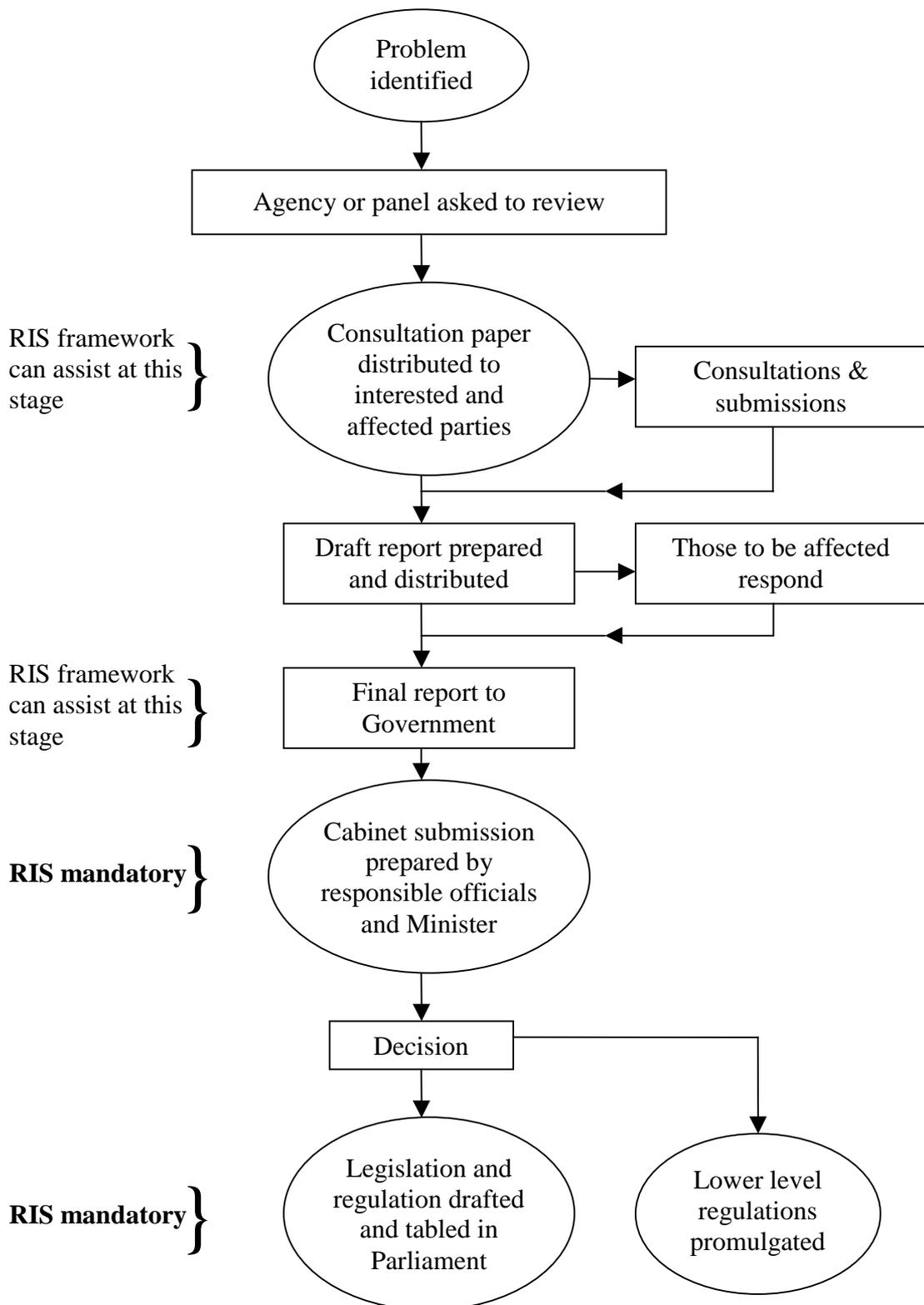
Box 3.4 Elements of a Regulation Impact Statement (RIS)

A RIS has seven key elements which set out:

- the problem or circumstances which give rise to the need for action;
- the desired objective(s);
- the options (regulatory and non-regulatory) that may constitute viable means for achieving the desired objective(s);
- an assessment of the impact (costs and benefits) on consumers, business, government and the community of each option (including the impact on small business paperwork and compliance costs);
- a consultation statement (the process and results of consultation);
- a recommended option; and
- a strategy to implement (including consideration of appropriate enforcement mechanisms) and review the preferred option.

Source: ORR (1998, p. 3).

Figure 3.2 **Stylised policy development process**



Box 3.5 **Comments on the RIS for an access regime for postal services**

Background

In July 1998 the Government decided on a range of further reforms to the postal services industry which included the introduction of an access regime to the Australia Post network. On 6 April 2000, the *Postal Services Legislation Amendment Bill 2000* was tabled in Parliament. The Explanatory Memorandum included a RIS that assessed three options for implementing an access regime.

Option A — Limited legislative intervention with no role for ACCC or equivalent arbitration.

Option B — Modified access regime based on Part IIIA of TPA with two components: Australia Post to submit an undertaking on terms and conditions for access and obtain ACCC acceptance; and ACCC to be given capacity to declare specific postal services.

Option C — Development of a comprehensive mandatory undertaking by the ACCC.

Assessment and conclusion

The RIS examined each of the three options, setting out the overall costs and benefits of each as well as those costs and benefits applying specifically to Australia Post, the remainder of the postal services industry and consumers.

A presentation of those facts led to the following conclusions:

Option A has the virtue of simplicity, but is unlikely to have the outcome of generating further competition in the postal services market to any substantial effect.

Option B meets the objectives for generating further competition in the postal services market, and having this competition flow through to the range of services offered to consumers, and the pricing of those services. It will ensure that the competition will develop as quickly as possible, both to address existing criticisms of the services available, and to provide a reasonable basis for assessing the effect of the reforms in the review in 2003.

Option B provides a balance by allowing for commercially negotiated outcomes, while providing a framework for arbitration by the ACCC, should this prove necessary. It is for this reason that it is the preferred outcome at this time.

Option C would be more suitable in a more stable environment where it was planned to operate over a decade or more.

In this RIS, the main good regulatory principles at issue were: further promotion of competition; allowing commercial negotiation to settle issues as much as possible; and, keeping compliance costs to reasonable levels.

3.8 Strengths and weaknesses of the RIS process

The most important and challenging facet of the RIS process is that it promotes a significant ‘cultural change’ among Commonwealth officials and the Government. It prompts abandonment of the traditional regulate-first approach, encourages assessment of non-regulatory options and seeks more careful selection and better justification of the preferred option. Changes of this nature are not achieved quickly, although, after three years, there is encouraging evidence that the RIS process is contributing to better regulatory outcomes.

An important strength of the RIS process is that it provides a common framework that can be used at all stages of the policy development process and can cope with a wide range of different problems and different regulatory agencies. Further, in contrast with several States which have Subordinate Legislation Acts that mandate RIS requirements, the Commonwealth RIS process is based on administrative decree.

Looking to the future, the RIS documentation will provide a valuable basis for a later comparison of impact assessments with actual outcomes. This analysis has the potential to better inform best practice in future regulatory design.

The RIS process makes the Government and its officials more accountable for policy choices and promotes improved transparency in policy making because the RIS documentation is made available to the public. The Government has instructed an independent agency, the Productivity Commission, to publish annually information on the extent to which Commonwealth departments and regulatory agencies satisfy the RIS requirements (PC 1999, Appendix B).

The RIS process gives those directly affected by the regulatory arrangements a very effective means of influencing the nature and effectiveness of the regulation, through a well-structured consultation process.

Weaknesses in the RIS process

For the RIS process to make a useful contribution, there needs to be an orderly policy development process (see figure 3.2) and the RIS needs to be done early enough to articulate viable options before positions become locked in. Quite often policy making does not fit this model. Rather, policy making often results from political imperatives calling for quick action, from deals with particular interest groups and from bargaining between political interests or parties. Sometimes these approaches result in good regulatory outcomes, but the risk of failing to do so is

undoubtedly higher than in cases where a more open and orderly policy development process is adopted.

Even then, there are situations when the RIS is prepared too late. Recently, a department provided the ORR with the nearly final draft of a Cabinet Submission and sought advice as to whether or not a RIS should be prepared. In such situations, the RIS becomes no more than an apology or rationalisation for the preferred option, in stark contrast to its intended role of setting out an even-handed description and assessment of all the viable options and their likely impacts.

A further weakness, it can be argued, is that the RIS process is not accompanied by adequate sanctions and penalties to ensure it is complied with and is effective. Sanctions include an unfavourable assessment going to Cabinet from the ORR, the Assistant Treasurer being informed of particular failures to comply and the publication by the Productivity Commission of failures to comply.

Much regulation making is incremental or evolutionary in nature, as modifications are made to improve or refine existing law. If the regulatory base is flawed, quality control processes focusing on changes to that base will have limited effect. Therefore, governments occasionally need to institute complete reviews of related sections of law, as the Commonwealth undertook with corporations law in recent years, and that such reviews utilise a systematic framework, such as that provided in a RIS.

With regard to quasi-regulation, because there is no central Commonwealth register of such regulations and many different processes for making such regulations, it is very difficult to track them down and even harder to impose a quality assurance process on them. For example, if an industry sector is persuaded by government to develop a code of practice, with some administrative and advisory support from government officials, it is not clear who will prepare the RIS and what will be done with it.

Finally, departments and regulatory agencies sometimes do not have sufficient skilled staff to prepare RISs. The ORR is responsible for training staff and providing advice and detailed feedback to agencies as they prepare particular RISs. That process does not always work as well as it should, because Commonwealth officials making regulations have widely different backgrounds in engineering, environmental sciences, nutrition, pharmaceutical chemistry and the law. Some of these officials have difficulty working with the analytical processes embodied in a RIS, though they typically are highly skilled in their own fields. In particular, the identification and assessment of the likely costs and benefits of different regulatory options can be a difficult task. While qualitative analysis is often sufficient for

informing decision making, in those cases where quantitative analysis is both appropriate and viable, it is seldom done well.

Current developments

The Government's mandatory requirement that RISs be prepared has now been in operation for three years. There is anecdotal evidence that the process is already contributing to better consideration and documentation of policy and regulatory options. However, the process is yet to yield its full potential. Three developments during the year 2000 will improve the outcomes delivered.

First, for the first time in May 2000, the Office of Small Business published nine regulatory performance indicators for each Commonwealth department and major regulatory agency (DEWRSB 2000, pp. 23–30). The indicators have been designed as a 'first step towards enabling benchmarking of government performance in regulation reform' (Prime Minister 1997, p. 80). Examples of some aggregate results for 1998-99 are:

- 88 per cent of RISs adequately addressed whether or not the proposal would provide a net benefit to the community;
- 39 per cent of regulations amended during the year incorporated flexibility;
- 91 per cent of agencies have communications strategies for regulation or have formal consultative channels for communicating information about regulation;
- 27 per cent of agencies publish an adequate forward plan for introduction and review of regulation; and
- 59 per cent of agencies have organisational guidelines outlining consultation processes, procedures and practices.

The publication of such results each year — a requirement set by the Government — will put pressure on all Commonwealth agencies to improve their regulation making process and, ultimately, to improve regulatory outcomes.

A second development that will contribute to better regulation making is the Government's requirement that, starting from the latter half of the year 2000, all Commonwealth agencies are to publish regulatory plans detailing information about regulatory changes in the previous financial year and those planned for the forthcoming year. The purpose of the annual regulatory plan 'is to bring a strategic focus to the regulation making activities of departments and agencies, and to give greater notice to the business sector of regulatory intentions' (DEWRSB 2000, p. 25). This initiative will put pressure on agencies to improve the transparency of

their regulation making and possibly to make greater use of the RIS process early in policy development.

Third, the Productivity Commission in its annual report, *Regulation and its Review*, is required to report publicly on compliance with the RIS requirements (PC 1998, 1999). By progressively increasing the amount of detail reported, the Commission is aiming to steadily improve the way agencies incorporate the ‘good regulatory principles’ outlined in this paper into the RIS process accompanying their regulatory proposals.

3.9 Conclusion

The Commonwealth Government has made a strong commitment to reducing the burden of regulation on business and improving the quality of regulation. Other Australian governments have in place broad strategies with similar objectives.

A principal means of achieving those objectives is to subject all regulatory proposals to a standard quality assurance process that helps ensure that ‘good principles’ are built into the regulation. The RIS process is such a mechanism which, when taken to its full potential, will lead to better — more effective and more efficient — regulatory outcomes.

References

- ACCC (Australian Competition and Consumer Commission) 1999, *Price Exploitation and the New Tax System, General Principles, Information and Guidelines on when Prices Contravene Section 75AU of the Trade Practices Act 1974*, ACCC, Canberra, March.
- Bridgman, P. and Davis, G. 2000, *The Australian Policy Handbook*, 2nd edn, Allen and Unwin, St Leonards, New South Wales.
- COAG (Council of Australian Governments) 1997, *Principles and Guidelines for National Standard Setting and Regulatory Action by Ministerial Councils and Standard-Setting Bodies*, November (amended).
- DEWRSB (Department of Employment, Workplace Relations and Small Business) 2000, *Annual Review of Small Business 1999*, Office of Small Business, DEWRSB, Canberra, May.
- Douglas, R. and Jones, M. 1999, *Administrative Law*, 3rd edn, Federation Press, Sydney.

-
- Government of Canada 2000, *The Regulatory Policy*, Privy Council Office, Ottawa, http://www.pco-bcp.gc.ca/raoics-srdc/regpol_e.htm (accessed 13 October 2000).
- Government of Victoria (undated), *Principles of Good Regulation*, Office of Regulation Reform, Melbourne.
- IDC (Interdepartmental Committee) on Quasi-regulation 1997, *Grey-Letter Law*, Commonwealth of Australia, Canberra, December.
- Kahn, A. 1989, *The Economics of Regulation — Principles and Institutions*, 2nd printing, MIT Press, Massachusetts.
- NCC (National Competition Council) 1997, *Legislation Review Compendium*, NCC, Melbourne, April.
- 1998, *Compendium of National Competition Policy Agreements*, 2nd edn, NCC, Melbourne, June.
- OECD (Organisation for Economic Cooperation and Development) 1997, *The OECD Report on Regulatory Reform*, OECD, Paris, June.
- ORR (Office of Regulation Review) 1998, *A Guide to Regulation*, 2nd edn, AusInfo, Canberra, December.
- PC (Productivity Commission) 1998, *Regulation and its Review 1997-98*, AusInfo, Canberra, November.
- 1999, *Regulation and its Review 1998-99*, AusInfo, Canberra, November.
- Prime Minister 1997, *More Time for Business*, Statement by the Prime Minister the Hon. John Howard MP, AGPS, Canberra, 24 March.
- Small Business Deregulation Task Force 1996a, *Working Overtime: A National Survey of the Paperwork Burden on Small Business*, Background paper no. 3, Commonwealth of Australia, September.
- 1996b, *Time for Business*, Commonwealth of Australia, November.
- Sturgess, G.L. 1994, *Thirty Different Governments ...*, Report of the Commission of Inquiry into Red Tape, Sydney, 31 January.

Discussant — *Aaditya Mattoo*

World Bank

Paul Coghlan's paper was both enjoyable and thought provoking. It is a little ironic that someone who has never had any direct contact with regulation should comment on a paper by someone who lives, breathes and writes regulation. I feel like the uneducated listener who upon hearing a Mozart symphony is supposed to have said 'Too many notes'. There are perhaps too many principles in this paper! I shall in my comments try to relate some of these principles to what we know about how regulations should work.

In assessing the appropriateness of regulation, the relevant questions are:

- What is the objective of the regulation?
- What is the best regulatory instrument to achieve the desired objective?
- How is the regulation best implemented?

Paul Coghlan's paper focused more on the third, and perhaps most important, question.

D3.1 When and how to intervene in markets?

The paper suggests two criteria — the probability of an adverse event and its impact (section 3.4) — to determine the appropriate form of regulation, ranging from no regulation through self- and quasi-regulation to 'black letter' law.

Note first that regulation may still be justified when the impact of an individual event is low, as in the supermarket-overpricing example, but the event happens on a systematic basis. The basic concern should be whether the market will produce an outcome superior to the one obtained under regulation. This is a relative choice, because the social optimum is not necessarily obtained through regulation, and regulatory failure is important, often more important than market failure. Once the possibility of regulatory failure is recognised, the existence of a market failure does not necessarily constitute a case for intervention.

To determine whether the market will deliver the desired outcome, it is necessary to look at the private incentives and how much of a need there is to modify them to achieve socially more desirable outcomes. The first step generally involves looking at the degree of competition in the market. But competition and incentives are not monotonically related. In some services, like telecommunications, the emergence of new technologies has resulted in more competition and the need for regulation has diminished. In other services, competition can create a free-rider problem. For

example, a free rider problem could arise in distribution of consumer durables. A retailer who provides costly but free pre-sales demonstration services may find that consumers inspect products in his shop and then buy more cheaply from a retailer who does not provide such services. In response to the problem of under-provision of such pre-sales services, countries have tended to be excessively permissive with regard to exclusive territories. The key point is that the regulatory approach depends on what the market will deliver in specific circumstances.

Sometimes concerns about reputation can exercise sufficient discipline on private behaviour. In services such as health, legal and accountancy services, information asymmetry is a problem, but the rewards to a good reputation can help to overcome the problem. It is also possible to remedy adverse selection and moral hazard problems through explicit and implicit contingent contracts. For instance, the possibility of being sued can exercise a significant discipline. In many cases, the government can significantly alleviate information asymmetries by directly disseminating information.

Each of the above options have their own limitations, but they need to be considered carefully in determining how much need there is for regulations such as qualification and licensing requirements. Ideally, the paper should have evaluated these options as alternatives to regulation more fully, perhaps in a sample of sectors. This would help determine, in terms of the paper, the probability of an adverse effect.

D3.2 Frameworks for regulating services

There is often scepticism towards a horizontal approach to regulating services. This is frequently based on the belief that services sectors are so different from one another that it is not possible to adopt a horizontal approach. This has been argued very strongly in a multilateral context but is also relevant in the national context in federal countries like Australia. However, the basic reasons for market failure or for regulation to achieve social objectives are not very different across service sectors. The economic case for regulation in all services sectors arises essentially from market failure attributable primarily to three kinds of problems, natural monopoly or oligopoly, asymmetric information and externalities.

The existence of a monopoly or oligopoly is the basic problem for a range of network services. Here, the *Telecommunications Reference Paper* (NGBT 1996) is an extremely powerful example of what can be achieved at the multilateral level. Perhaps one of the most significant achievements of the GATS is the establishment of a simple core principle — incumbents must provide cost-based access to essential facilities. While people will say ‘how do you determine what is cost-

based?', the principle at least provides a framework for challenging more egregious departures from what might be cost-based.

The other major problem in services is that of asymmetric information. In this area, it is not the market failure that is so much at issue in the trade context. Rather, local measures to correct the market failure, through qualification requirements or technical standards, can themselves become impediments to trade. How is the protectionist measure to be distinguished from the legitimate? The legal 'necessity test' and the economic 'theory of domestic distortions' can be brought together very nicely. Is the measure necessary can be interpreted to mean is the measure the best way of addressing the market failure. I believe that the alternative interpretation to be found in WTO jurisprudence 'is this the least *trade restrictive* measure' is politically unsustainable. Regulators might legitimately respond that 'this is the best instrument. I don't care whether or not it is trade restrictive'.

How would the approach I suggest work? For example, in the case of qualification requirements, if the problem is the inability to see whether a doctor is adequately trained or not, the way to correct that problem is not through excessively burdensome requalification requirements, but through a test of competence. How far can we go towards creating a presumption in favour of a test of competence? This in turn depends on the extent of country-specific versus universal knowledge legitimately required of a professional. For instance, while there might be a high universal content in what a doctor is required to do, since a human body is a human body, one legal or tax system is not the same as another. Still requalification and retraining must not be more than is necessary. Thus, the necessity test is a form of forced recognition — you cannot deny people the right to be recognised as professionals if they are adequately trained.

D3.3 Measuring the regulatory barriers to services

The World Bank is keen to forge closer collaborative links with researchers throughout the world undertaking services-related research. Apart from the standard-specific work, the World Bank is also undertaking research to measure the barriers within specific service industries. One approach we are considering is whether it is possible to measure the gap between actual and the economically efficient regulation. This work seeks to build on the extremely valuable work that has been done at the Productivity Commission.

Reference

NGBT (Negotiating Group on Basic Telecommunications) 1996, 'Reference paper on telecommunication regulation', WTO, Geneva.

General discussion

The discussion focused on:

- changing business attitudes to regulation; and
- regulatory failure.

Changing business attitudes to regulation

One participant said that the most important contribution that the *Guide to Regulation* had made was to alter the climate of opinion in the business community about regulation. It had got business thinking about no regulation or self-regulation rather than ‘black letter law’. An example was the review of the *Australian Food Standards Code* that changed the way that food and wine businesses thought about what regulation could do for them. The role of the WTO was also important because the view had often been about particular industry concerns, forgetting that Australia also had obligations abroad. If good practices were achieved at home, Australia had the moral high ground to push for more access abroad, in particular removing technical barriers to trade.

An important part of information dissemination was changing the attitude of business to regulation. Contrary to common perceptions, business was actually quite keen on regulation. It was important that business demonstrate that regulatory proposals result in net benefits to the community as a whole, not just to the industries concerned.

Information dissemination was important and used a lot by government (eg anti-smoking and road safety campaigns). Information dissemination would fall into the bottom left hand corner of the regulatory spectrum in figure 3.1, but it was not strictly a regulatory measure.

Regulatory failure

Another participant suggested that Paul Coghlan’s paper be expanded to talk about the risk of regulatory failure, which was really where the tradeoffs lay. In addition, while rapid technological changes can cause problems in keeping ‘black letter law’ up to date, they also meant that market failures were likely to be transitory and the scope for regulatory failure was dramatically increased.

In response, Paul Coghlan agreed with Aaditya Mattoo’s point about systemic low-impact problems. In such cases, as in the supermarket example, self-regulation may not address the problem. Figure 3.1 was trying to say that looking at regulatory

forms was not enough, but that a cost-benefit assessment of a range of options was needed. If done properly, a RIS should do this. The first step involved defining the problem that the regulation sought to address. In doing so, the RIS should articulate whether the problem was random or systemic, how big the problem was and so forth. The RIS should pick up the sort of low impact, systemic problems that Aaditya Mattoo identified.

Regulatory failure was apparent and endemic. A role of the ORR was to take the long-term approach of promoting a quality assurance process to regulation. The Productivity Commission's inquiries also identified key regulatory failures after they occurred.

PART C

REGULATING FOR
COMPETITION

4 Access: what, where and how

Stephen King

University of Melbourne

4.1 Introduction

It is approximately seven years since the Hilmer report on National Competition Policy (Hilmer 1993) brought infrastructure access to the centre of the Australian regulatory stage. Since then, the Commonwealth and all State governments have introduced access regimes. There are general access provisions under Part IIIA of the *Trade Practices Act 1974* (Cwlth), as well as specific regimes for telecommunications, rail, airports and other infrastructure industries. The administration of these new rules is vested in a variety of regulatory bodies that were unknown seven years ago — the Australian Competition and Consumer Commission (ACCC), the Office of the Regulator General (ORG) and the Queensland Competition Authority to name but a few.

Infrastructure access has also dominated regulatory reform in energy, telecommunications and transport internationally. Many of the issues that are currently being faced in Australia are also being tackled in the United States, the European Union and Asia.

This paper provides an overview of infrastructure access, and considers how the access problem might be solved in certain circumstances. Along the way, it provides a brief critique of some elements of the current Australian access laws under Part IIIA of the Trade Practices Act. Finally, some of the ideas developed below are then applied to rail access.

4.2 Why is there a problem?

Access is only an issue when one or more firms each control upstream facilities that provide a good or service that is needed for further downstream production. Other firms, which seek to enter the downstream market, either have to purchase the

relevant input from one of the firms, or integrate back into upstream production themselves.

This situation characterises almost any vertical production chain. But this does not normally lead to concern. If there are a reasonable number of non-integrated upstream firms, each of which can provide the relevant input, then competition will ensure that potential downstream firms can buy the input at a reasonable price. If the upstream firms are all integrated into downstream production and are competitive, then the vertical production chain as a whole will lead to an economically efficient outcome. Any inability of a potential downstream firm to purchase the upstream input is likely to reflect that such sales would be costly and inefficient, rather than an anticompetitive intent on the part of an integrated firm. If a potential downstream competitor was unable to purchase the upstream input, it could enter upstream production itself. Any attempt to prevent downstream competition by withholding the upstream input would simply invite competition over the entire production chain. Finally, even if there is only a small number of integrated firms, each of which refuses to sell the input, and further integrated entry is not possible, there is still only a problem if there are no other substitute downstream products that act as competitive constraints on the integrated firms.

For access to be a problem, the vertical production chain must be economically dysfunctional. There must be little or no competition in the provision of the upstream good or service. There must be little or no competition in the downstream market from alternative products that do not require the relevant input. And it must be extremely difficult or economically undesirable for any potential entrant to enter upstream production. In these circumstances, a potential downstream entrant may either be refused permission to buy the relevant input, or the input may be priced at a level that militates against effective competition (Hilmer 1993, chapter 11).

Why refuse to sell the input?

For simplicity, suppose that an integrated monopoly controls the relevant upstream production. Why would this firm ever seek to refuse sales of its upstream product to potential downstream entrants? While such sales would invite downstream competition, it is far from obvious that this is undesirable for the upstream firm. So long as it can set an appropriate (nonlinear) access price, the integrated firm may be able to retain all monopoly profits from the vertical chain of production and seize all of these profits through input pricing.

To see this, suppose that one unit of the upstream input is required for one unit of the downstream output. Let the initial monopoly downstream price be denoted by P^m . Suppose that there are n potential entrants to downstream production. If these

firms buy the upstream input at price p_a and compete in the downstream market with the integrated incumbent, then the downstream price will be $P(p_a, n)$. It is reasonable to assume that P is continuous in p_a , $\left(\frac{\partial P}{\partial p_a}\right) > 0$ for any n , if p_a is set equal to the marginal cost of upstream production then $P \leq P^m$, and if p_a is set sufficiently high then $P \geq P^m$. Given these assumptions, the integrated firm can always set an access price such that downstream competition results in exactly the integrated monopoly price.¹ The integrated firm can then seize any profits from the downstream competitors using a fixed fee (such as an annual ‘access charge’). Overall, the integrated firm retains all monopoly profits, despite allowing downstream competition.

If an integrated firm refuses to supply access to potential downstream competitors, then this probably reflects the failure of one or more of the assumptions given above. The monopolist might be unable to set the correct nonlinear price, possibly because it is uncertain about the exact nature of potential downstream competition or because it is unable to credibly commit not to renegotiate nonlinear access prices with some downstream competitors.² Downstream production may not be equally efficient as integrated production. For example, if there are substitution possibilities between the monopoly input and some other input, then any attempt to manipulate the upstream price might lead to an inefficient mix of inputs and a diminution of total profits. Alternatively, there might be real economic costs of providing the upstream input.³

If an integrated monopolist refuses to supply an upstream input, then this reflects that it is either economically inefficient to supply the input while maintaining monopoly output or there are limitations on contracts that prevent the upstream firm from maintaining monopoly output downstream.

A regulatory solution to this ‘refusal to deal’, such as, say, simply requiring that the relevant input be sold, may lower social welfare. For example, if the integrated firm does not wish to sell access because this leads to inefficient production, then

¹ This will often be *easier* than it sounds. Essentially, the upstream firm can control downstream production indirectly through a contract that sets the amount of the upstream product each downstream firm can purchase. This would simply be a contract where p_a is very low up to a certain quantity of the input, then infinitely high for larger input quantities.

² McAfee and Schwartz (1994) and Rey and Tirole (1996) consider this issue for a non-integrated supplier of an upstream input.

³ Of course, if potential downstream firms are more efficient than the integrated firm or are able to sell differentiated products that raise total profit above the integrated monopoly level, then the integrated firm will want to sell the upstream input. Allowing competition will raise the profits that it can seize through nonlinear pricing.

mandating that the input be sold may simply result in higher costs and lower social welfare. The integrated firm will set a price for access that maintains maximum profits and there will be ‘competition’ downstream. But to the degree that marginal production costs rise, final product consumers will tend to face a higher price, not a lower price.

Alternatively, if the integrated firm initially refused to supply access due to difficulties in maintaining the profit maximising price, then mandating access is likely to result in lower final product prices, raising social welfare.

Is there a problem if the input is being supplied?

If an upstream monopoly (whether integrated or not) does sell the input product to downstream firms, then it will want to price the input in a way that:

- maintains overall monopoly production and profits downstream; and
- enables the upstream firm to seize as much of the monopoly profits as possible.

In either case, sale of the essential input will not promote efficient downstream pricing, although it creates the appearance of downstream competition. The real problem in these circumstances is the lack of upstream competition. Such a problem is not, however, unique to input markets.

There are numerous markets in every economy where it might be felt that competition is below some desirable level. However, it is not necessarily desirable to immediately intervene in such markets through, say, price controls. Regulation has its own costs and if any lack of competition is transitory, then it is likely that these costs will outweigh any benefit. Further, regulation might exacerbate competitive problems. If an industry is subject to intrusive regulation, then this may prevent firms from making profits that will attract new entrants. In many markets, it is better to allow competition to develop naturally.

This also holds for most input markets. If there is a current lack of competition, then it is likely that this will diminish over time. While the rules of Part IV of the Trade Practices Act need to be rigorously enforced to make sure that increased competition is not artificially impeded, there is little to be gained from direct regulation.

In this sense, access to an essential input is only a cause for specific regulatory intervention if it is either highly unlikely that competition will develop in the upstream market in the longer term or if such competition is itself undesirable. In general, this means that the upstream production process involves a natural monopoly technology.

A natural monopoly technology exists if, at all relevant levels of output, it is more efficient (in terms of lower production costs) to have the output supplied by a single firm than by more than one producer (Panzar 1989, Waterson 1987). In such circumstances, it is economically undesirable for there to be competitive supply in the sense that, *given a level of output*, the costs of having two or more firms produce that output is higher than if only a single firm produced that output. The existence of a natural monopoly technology is an empirical question, although such technologies are most likely to arise where production involves fixed costs that are large when compared with marginal costs. Further, the existence of a natural monopoly technology depends on the extent of demand. As demand for a product grows, it may become economically desirable to divide the increased production between different producers.

If production involves a natural monopoly technology, then this does not mean that competition in production is impossible. Rather, it means that this competition is inefficient in the sense that total output is not produced at minimum cost. Whether competition between two or more firms can arise for a product that involves a natural monopoly technology depends on the nature of interactions between the relevant firms. In particular, such competition, if it exists in the longer term, is likely to be muted.

Competition with a natural monopoly technology is an extreme version of a common economic phenomenon. If production involves fixed costs and there is imperfect competition, then free entry by firms will not lead to cost efficient production in the sense that, *given the level of output*, total cost could be reduced if there were fewer firms in the industry (Mankiw and Whinston 1986). Such common inefficiency does not call for general regulatory intervention, because of the costs associated with regulation. Limiting entry might reduce costs for a given level of production, but is more likely to simply lead to a reduction in output, a rise in prices and a loss of allocative efficiency that more than offsets any productive gain. Imposing price regulation on the industry is likely to lead to greater inefficiencies, as firms inflate costs to try and convince regulators to raise prices. Similarly, if there is a natural monopoly technology in upstream production, any benefits that may arise from an access regime need to be weighed against the costs of regulation.

Summary

The discussion above highlights five key points:

- access is only an issue if the relevant upstream input is essential and the vertical chain of production is competitively dysfunctional;

-
- access regulation (as opposed to other regulatory solutions) is only likely to be needed if upstream production involves a natural monopoly technology;
 - existing sales of the input by an upstream monopoly, whether integrated or not, do not mean that there is not an access problem;
 - light-handed regulation that simply requires (negotiated) sale of the relevant upstream input may or may not raise social welfare; and
 - the costs of regulation need to be weighed against the potential benefits, particularly where there is current (albeit inefficient) upstream competition.

4.3 The legislative solution to the access problem

The current access procedures in the Trade Practices Act attempt to both define the access problem and provide a regulatory process to overcome this problem. Part IIIA provides the general access regime and begins with the *declaration* of a service by a relevant government minister, following a recommendation by the National Competition Council (NCC). Section 44B states that ‘service means a service provided by a facility and includes:

- (a) the use of an infrastructure facility such as a road or a railway line;
- (b) handling or transporting things such as goods or people; ...

but does not include:

- (f) the use of a production process; except to the extent that it is an integral but subsidiary part of the service.’

Under section s. 44G2 of the Trade Practices Act, the ‘Council cannot recommend that a service be declared unless it is satisfied ...

- (a) that access (or increased access) to the service would promote competition in at least one market (whether or not in Australia), other than the market for the service; and
- (b) it is uneconomic for anyone to develop another facility to provide the service’.

If a service is declared, then the facility owner and the access seeker(s) negotiate an access agreement. If negotiations are unsuccessful, then either an access seeker or the facility owner can apply to the ACCC to determine the terms and conditions of access.

The Trade Practices Act provides two alternatives to declaration under Part IIIA for access. First, if the ACCC accepts an access undertaking from a facility owner

under s. 44ZZA of the Act, then the service cannot be declared (s. 44H3). An undertaking must include terms and conditions of access and can only be accepted by the ACCC prior to declaration (s. 44ZZB). Second, a service cannot be declared if it is already subject to an effective access regime (s. 44G2e), such as a State-based access regime, established in accordance with the Competition Principles Agreement.

The access rules in the Trade Practices Act, and particularly those highlighted above, have been subject to extensive discussion (eg King 1997a; King and Maddock 1996a, 1999a; NCC 1996). It is far from clear that the access rules capture the true nature of essential facilities. Further, by focusing on negotiation rather than regulation, the access process established by the Act fails to recognise the potential for simple ‘rent sharing’ between the access provider and access seekers (King and Maddock 1996a, 1999b). Finally, the undertaking route ignores the potential desirability of *not* allowing access. Each of these issues are now considered in turn.

Declaration and essential facilities

Before considering rules for declaring services produced by essential facilities, it is necessary to formally define such facilities. An essential facility involves two distinct characteristics (King and Maddock 1996a, 1996b). First, it must involve a natural monopoly technology. This is formally defined above. Second, the product that the facility produces must be an input into the production of a specific final good or service where:

- there is no alternative input or process that enables another firm to produce an equivalent final good or service at a comparable or lower cost; and
- there is no alternative substitute final good or service that can be manufactured and sold at a comparable price without using the input product.

The existence of a natural monopoly technology means that competition in upstream production does not minimise production costs. Further, where such competition exists in the long term, it is likely to be muted. However, this does not mean that there is a potential for monopoly pricing of the input, unless there is neither effective competition from alternative inputs nor from alternative final products.

The tests under s. 44G2 of the Trade Practices Act only partially reflect these essential facility characteristics. The requirement that access promote competition in another market can be viewed as an attempt to capture only inputs and to focus on dysfunctional vertical production chains. However, the market interpretation used by the NCC when considering this test has not resulted in a focus on final

goods and services, but rather artificial ‘functional’ distinctions in a production chain (eg NCC 1997). This focus on competition at other intermediate steps of a vertical production chain hides the fact that access is only useful if it results in more efficient production and pricing of final products.

The requirement that it is uneconomic for anyone to develop another facility to provide the service may be thought of as a natural monopoly test (NCC 1996). However, if ‘uneconomic’ is interpreted as meaning that it is not profitable for a private investor to develop another facility, then the test clearly fails to isolate facilities with a natural monopoly technology. As noted above, it is possible for more than one facility to exist even with a natural monopoly technology. Further, if it is privately unprofitable to develop another facility, then this simply means that the upstream market is ‘saturated’. This can arise under perfect competition.⁴

The NCC has considered an alternative interpretation where ‘uneconomic’ means that it is socially undesirable to develop another facility. This would involve an explicit natural monopoly test, but it is far from clear that the current wording of the Trade Practices Act allows for this interpretation.⁵ Further, determining precisely what is and what is not a natural monopoly technology can be difficult.

A pragmatic approach to essential facility declaration would focus on the two defining characteristics of essential facilities and try to capture these characteristics in a workable test. It would also recognise the potential for over-regulation and the costs of legal conflict. As such, the rules would tend to be under-inclusive, rather than over-inclusive.

As an example, consider the following tests that could replace ss. 44G2a and b.

The Council cannot recommend that a service be declared unless it is satisfied that:

- (a) access (or increased access) to the service will substantially increase competition in the market for a final good or service;

⁴ In fact, the standard textbook version of perfect competition defines long-run equilibrium as the situation where it is ‘uneconomic’ for any firm that is not currently in the industry to ‘develop’ its own facilities and enter the industry.

⁵ The Australian Competition Tribunal, in its recent decision regarding Sydney International Airport (1 March 2000), affirmed the social interpretation: ‘The Tribunal considers, however, that the uneconomical to develop test should be construed in terms of the associated costs and benefits of development for society as a whole. ... If ‘uneconomical’ is interpreted in a private sense then the practical effect would often be to frustrate the underlying intent of the Act’ (para 204-205) .

-
- (b) the service is an input into further production and there is no alternative service or production process available to the access seeker that reasonably can be used for the production of the relevant final good or service; and
 - (c) efficient production of the service necessarily involves an infrastructure facility with high fixed costs and relatively low operating costs so that it is likely that the development of another such infrastructure facility by any person will raise the total cost of supplying the market for the relevant final good or service.

Rule (a) focuses on final markets for goods and services. The requirement that there is a substantial increase in competition guarantees that there are no relevant substitute final goods or services that can be produced without the declared service.

Rule (b) focuses on input services and guarantees that there are no reasonable alternative inputs. Importantly, this rule requires that there is no alternative service. Clearly, such an alternative would be available if there were two separate facilities that could provide the service. As such, rule (b) requires that there is only a single producer of the service. As noted above, it is possible for there to be an essential facility problem even with more than one input producer. However, it is necessary to weigh up the regulatory costs against the potential for competition. Two firms do not necessarily guarantee competition. But it is necessary to draw a line somewhere and if there are two potential providers of the essential input then the need for intrusive regulation through declaration is significantly diminished.

Rule (c) provides an explicit natural monopoly technology test. It focuses on facilities with high fixed costs, as these are most likely to involve natural monopoly technology. The word ‘likely’ is included to avoid excessive debate about whether the facility does or does not involve a natural monopoly technology.

Pricing by negotiation or regulation

Under the current access regime, declaration leads to negotiation. However, there is no reason to believe that access negotiations will lead to economically efficient pricing. Rather, access seekers and the access provider have incentives to try and formulate access prices that both maintain monopoly profits and share those profits between themselves.

The current regime is more ‘light-handed’ than the original proposal presented by the Hilmer Committee, which involved the NCC providing access pricing principles when a service is declared (Hilmer 1993, pp. 266–7). While the provision of such principles, or any other rules that may effect a potential ACCC determination over an access dispute, can affect the outcome of access negotiations, they do not prevent

the parties agreeing on anticompetitive prices (King and Maddock 1999b). Because of this, any declaration needs to be accompanied by a regulatory process of setting access prices. At a minimum, either the NCC or ACCC need to formulate a set of access pricing principles. After declaration, access seekers and the access provider would then negotiate access prices subject to satisfying these principles. Any agreement between the negotiating parties would need to be ratified by the NCC or the ACCC before coming into force. The issue of access pricing is discussed in more detail below.

The Part IIIA access regime, even with regulated post-declaration pricing, is unlikely to capture essential facilities where access is already being supplied but at a monopoly price. There is, however, no rule against monopoly pricing in Australia. Thus, if access is being provided at a monopoly price, then this is just a specific example of a more general limitation of Australian competition laws.

Undertakings and new investment

The potential for declaration can significantly impede new investment in facilities that embody a natural monopoly technology.⁶ To see this, suppose that there is a single firm that can build a new infrastructure facility that might be subject to declaration. For example, the facility might be a cable network in a country town that is to be used for pay-TV distribution. When the firm invests in this facility it faces an uncertain payoff. For example, suppose the network will cost \$51 million to install, but the firm is unsure of the size of demand for pay-TV in the town. If the pay-TV service is successful, then the firm will be able to set a relatively high price and it will receive (in present value terms) \$100 million. If the pay-TV service is moderately successful, then the firm will set a lower price and receive \$60 million. If the pay-TV service is unsuccessful, then the firm receives only \$20 million. Suppose that the probabilities of each of these outcomes are 25 per cent, 50 per cent and 25 per cent respectively. Then the expected return to the investor is \$60 million, less the \$51 million cost of the investment. So the firm expects to make \$9 million on average and (if risk neutral) will invest in the cable network.

Now, suppose that after building the network, the pay-TV distribution services provided by the network could be declared. A potential access seeker can wait until they have observed whether or not pay-TV is successful before they attempt to gain access. If declaration is successful, the network owner will have to sell access to at least one other pay-TV provider. If the network owner and access seeker

⁶ Gans and Williams (1999) and Gans (1998) note that when different firms can build an 'essential' facility then the principles used to set access prices can either create a race to be the first to invest or a situation where both firms want to wait for the other firm to build the facility.

successfully negotiate on access and this leads to no diminution in total profits, then the network owner will still need to share these profits with the access seeker. Alternatively, if access negotiations fail and the ACCC establishes access prices, then these prices are likely to encourage competition and reduce total profits below the monopoly level. Either way, when the network is declared, the network owner will receive lower total profits.

Using the numbers above, suppose that it is only worthwhile for another pay-TV provider to seek declaration if demand is high in the town and, in the absence of declaration, the incumbent would receive \$100 million. Further, suppose that in this situation declaration occurs and, due to access, the return to the network owner is reduced to \$60 million. *Ex post* this leaves the network owner receiving \$9 million more than the cost of the network. But *ex ante*, the investment in the network is no longer viable. The network owner now expects to receive \$60 million if pay-TV is successful, \$60 million if pay-TV is only moderately successful and \$20 million if pay-TV is unsuccessful. Using the same probabilities as before, this gives an expected return of \$50 million, less than the capital investment of \$51 million. The expected return from the investment is now negative \$1 million dollars. Hence, if the network owner believes that its network will be declared if pay-TV is successful, then they will not bother to invest in the network in the first place.

This example illustrates a general principle. Declaration and access are *ex post* decisions that affect the expected return from an investment. Further, access is most likely to be sought when the downstream returns from the investment are high. This means that, *ex ante*, the potential for declaration will ‘cream skim’ the returns from an investment and might make socially desirable investment privately unprofitable.

The principle, that access will deter investment when returns are uncertain, holds regardless of the access prices, so long as access leads to some diminution in total profits to the access provider. It is hard to imagine non-trivial access prices where this would not occur. The efficient component pricing rule (ECPR) for access would enable the access provider to retain all of its profits and so would not reduce investment. But, in the absence of final market price regulation, this rule is essentially trivial. Some authors have attempted to extend the rule to avoid triviality. For example, M-ECPR uses the market prices of substitutable final products as a basis for determining access prices (Economides 1997). However, if these final products are really relevant substitutes, then there is no essential facility problem and access is unnecessary. However, any lower access prices will reduce the returns from infrastructure investment and, as such, may make marginal but socially desirable investment, unviable.

Access pricing rules that allow a ‘reasonable’ return on investment do not avoid this problem. For example, suppose that if declaration is successful, access prices are set

by the regulator to cover the cost of the investment, including a ‘risk premium’. So long as there are some potential situations where the investment will be *ex post* unprofitable, the potential for access will distort the expected investment returns and may make the investment unprofitable. In the above example, the network owner did receive a ‘reasonable’ return on the investment even after access. However, *ex ante* this return did not offset the potential loss that could be incurred by the investor.

Even if the regulator could set the access price *before* any investment, the potential for access to distort investment would not be eliminated. To see this, suppose that the regulator can *ex ante* set the access price such that whenever access is sold the network owner is guaranteed to make enough return to cover their investment. Then this will still not overcome the problem if there is a potential for the service to ‘fail’. In these situations, there will be no access seeker and the investor will be forced to bear the entire burden of any loss. In the example above, the regulator could *ex ante* require that the investor receive \$60 million whenever access is provided, but as we have seen this will still make the project unviable as it does not protect the investor from the potential of a \$31 million loss if pay-TV is unsuccessful.

In summary, whenever the returns from a large infrastructure investment are uncertain, the potential for declaration and access (at non-trivial prices) will tend to deter socially desirable investment.

The access regime in the Trade Practices Act does not provide any way for new investors to protect themselves from declaration. At best, an investor can provide an undertaking to the ACCC that establishes ECPR as the rule for access. However, it is not clear that the ACCC would accept such an undertaking.

To avoid the disincentives for investment created by declaration, investors in new infrastructure facilities should be allowed to apply for an ‘access holiday’. In other words, investors should be allowed to present the ACCC with an undertaking such that the terms and conditions for access are that the infrastructure owner will not provide access for a fixed period of time, such as twenty years. The ACCC should be able to accept such undertakings when there is a reasonably high degree of uncertainty surrounding the investment, such that (say with a probability of at least 20 per cent) the investment might fail.

4.4 Access pricing

If a service is declared and a regulator is required to set access prices, what principles should be used to determine those prices? The problem of access pricing has led to considerable debate in the economics literature. This section provides an

elementary introduction to the standard economic theory of access pricing. Where relevant, references to more comprehensive or advanced material are provided.

Pricing and allocative efficiency

A simple economic principle underlies access pricing that promotes allocative efficiency. In the absence of any offsetting effects, the optimal price is set equal to the short-run marginal cost (SRMC) of production for the last unit sold (Doyle and Maher 1992, Kahn 1988, Slater 1989, King and Maddock 1996a).

The SRMC of production is the extra cost associated with a one-unit increase in output. For example, if a firm needs to increase its use of labour by one hour at a wage of \$50 per hour and needs to buy additional raw materials at a cost of \$100 to produce an extra unit of output, then the SRMC of that extra unit is simply the total cost of the additional inputs, \$150. If production occurs in a factory that is already rented by the firm on an annual contract then the SRMC would not include any part of that rent. The rent is a fixed cost for the firm in the short run. The SRMC only includes those costs directly associated with the additional unit of output.

SRMC is based on economic, rather than accounting, costs. For example, consider that in addition to an extra hour of labour and the additional raw materials, a one-unit increase in production requires the use of a machine that could otherwise be used to produce an alternative product. Then any (economic) profits that would have accrued from producing that alternative product represent an opportunity cost to the firm. Such opportunity costs need to be included in SRMC.⁷ Similarly, if the firm increases its production and this leads to a social loss that is not borne by the firm, this loss should be considered part of SRMC. An example is the social cost of pollution.

If the relevant production facility is capacity constrained, so that no extra output can be produced in the short run, then the SRMC is adjusted to allow for congestion. In this situation, efficient pricing involves rationing the infrastructure service on the basis of price.

⁷ While the opportunity cost of *inputs* is part of SRMC, the opportunity cost of the *output* is not part of SRMC. For example, consider that access can be used by the facility owner in its own downstream production or can be sold to a downstream competitor. Then the foregone profits from own downstream production when a unit of access is sold to a competitor are an opportunity cost of that sale, but are not a cost of providing access. This failure to distinguish between an opportunity cost of production and an opportunity cost from alternative disposal is incorporated into the standard definition of the efficient component pricing rule. See, for example, Baumol and Sidak (1994, p. 94).

SRMC pricing promotes allocative efficiency because it sends the correct economic signals to anyone wishing to obtain the relevant good or service. From an economic perspective, an extra unit should only be produced and sold if the value to the person purchasing that unit is at least as great as the cost of providing that unit. If price were below SRMC, there would be excessive demand for the product. If the producer were to meet this demand, then some consumers would purchase the good, despite the marginal cost of production being greater than their personal value. Such consumption would be socially wasteful.

It is also undesirable to set the price so that a potential purchaser who values the output at more than the marginal cost of production is dissuaded from buying the product. Such a situation would lead to a social loss in welfare, as the cost of producing the product is less than its value to the purchaser. The optimal regime involves price equal to SRMC.

While SRMC pricing maximises allocative efficiency, it presents two practical problems that impinge on productive efficiency and incentives for efficient investment. First, if the regulator requires access to be offered at SRMC, then the facility owner may have little incentive to reduce this cost. Any effort taken by the owner that improves productive efficiency by lowering SRMC is transferred into lower access prices. While time lags between the realisation of cost savings and the adjustment of access prices by the regulator may provide the facility owner with some temporary return from improved productive efficiency, the owner is unlikely to face the optimal incentives to engage in cost minimising activities.⁸

Second, access services that are produced using large infrastructure facilities such as transmission grids, rail systems or networks of pipelines, often involve significant returns to scale. Average production costs tend to fall for these facilities as output rises until the facility becomes congested, and SRMCs are generally less than average production costs. SRMC pricing does not allow the facility owner to cover his costs unless the facility is so congested that the congestion rents under SRMC pricing are able to cover the owner's average costs. A facility owner who believes that they will be forced to price access at SRMC will either refuse to build a facility or will so restrict the capacity of the facility that the congestion rents will allow them to make at least an adequate return on their investment. Neither situation is likely to be economically optimal.

⁸ Laffont and Tirole (1993) examine this problem and potential regulatory solutions in detail.

Additions and alternatives to SRMC pricing

Non-linear pricing

While SRMC pricing requires that the marginal price for access is set equal to SRMC, it does not require that the infra-marginal price is also set equal to SRMC. By allowing a facility owner to use nonlinear pricing schemes it is possible to generate a reasonable return on investment without constraining capacity in an undesirable way.

A two-part tariff, which involves an upfront fee and a per unit price, is the simplest form of nonlinear pricing scheme. By setting the per unit price equal to SRMC, a two-part tariff retains allocative efficiency.⁹ At the same time, the upfront fee can be used to cover the costs of infrastructure development. Freebairn and Trace (1992) recommend a two-part tariff for pricing railway services to coal producers:

The first-part tariff ... would cover each mine's allocation of unattributable costs as well as the capital costs of dedicated infrastructure The second-part tariff would be a per tonne of product charge based on marginal costs. (p. 37)

While the per unit charge of a two-part tariff should be set at SRMC to minimise allocative efficiency, the upfront fee can be used to promote efficient investment incentives (eg Gans and Williams 1999, Gans 1998).

When setting the upfront part of a two-part tariff, it may be desirable to allow the access provider to discriminate between access purchasers. Setting a uniform upfront fee, say, at a simple average of facility development costs, may exclude some low value users from purchasing access, even though these consumers have a marginal willingness to pay that exceeds SRMC. Allowing the access provider to set a lower upfront fee for these users, but a larger fee for higher value access seekers, will promote efficient facility use.

Price discrimination will only be possible if resale between access seekers can be prevented. Also, both efficiency and equity considerations will require a limit to be placed on the profits the access provider can make from the upfront fees. Without such a constraint, the access provider will have an incentive to raise upfront fees. This will provide two benefits to the access provider. Raising the fee will limit the degree of downstream competition by creating a barrier to entry for new firms. The high fee will also enable the access provider to reap most of the profits that are created by restricting downstream competition.

⁹ A limitation on the ability of nonlinear prices, such as two-part tariffs, to maintain allocative efficiency arises when downstream firms that purchase access as an input for production cannot themselves pass nonlinear prices on to final customers. See King (1997b) for a discussion.

While two-part tariffs are the simplest form of nonlinear prices, it is possible to design far more complex tariffs. Common examples include ‘block tariffs’ and quantity discounts.¹⁰

Rate of return regulation

The use of nonlinear tariffs can allow the infrastructure owner to gain a reasonable return on investment. However, it may be difficult for the regulator to judge whether the prices set by a facility owner represent a reasonable return or abuse of market power. To the degree that the facility owner is able to reap monopoly profits through a nonlinear access pricing scheme, and in so doing distort final product competition and prices, the regulator will want to control those profits. A traditional approach to this problem involves valuing the infrastructure owner’s capital stock to form a ‘rate base’ and allowing the owner to set access prices that generate no more than a regulated return on this base. This regulation, which has been used extensively in the United States for more than a century, is called rate-of-return (ROR) regulation.

While ROR regulation is often linked with uniform pricing (above SRMC), this need not occur. Given the allowed return on the rate base, the facility owner should be encouraged to set prices that lead to optimal facility use.

However, ROR regulation presents other problems. Like other regulations that limit the facility owner’s profits, ROR regulation reduces the incentives for the owner to minimise production costs. In fact, ROR regulation creates incentives for the access provider to deliberately choose an inefficient mix of productive inputs. If the access provider expands capital inputs, then this will increase the rate base and raise allowed profits. Consequently, ROR regulation tends to be associated with over capitalisation and over investment in infrastructure facilities.¹¹ While solving the problem of allowing the owner to receive a reasonable return on his investment, ROR regulation can exacerbate productive distortions.

Ramsey pricing

In some situations, it might be impractical to use nonlinear pricing for access. In such cases, it may be necessary to raise per-unit prices above SRMC to provide the

¹⁰ Mitchell and Vogelsang (1991), especially chapter 5, provide a more detailed analysis of a variety of nonlinear pricing schemes.

¹¹ This is sometimes referred to as the Averch-Johnson effect. It was first shown formally in Averch and Johnson (1962). For a discussion of ROR regulation, see chapter 21 in Carlton and Perloff (1994).

facility owner with a reasonable return on his investment, while at the same time seeking to minimise the allocative loss created by the access prices. The standard approach to this problem involves pricing to different customer groups according to the relative responsiveness of their demand. For example, if there are two distinct types of firms that purchase access, and only one type can (inefficiently) choose an alternative input, then it may be socially desirable to set a higher per-unit access price to those firms that have no alternative input.¹² Allocative distortions are created by changes in the quantity of a good that is used. So if one type of firm has more limited input substitution possibilities, then raising their access price will lead to a relatively smaller distortion in the purchases of the access input.

Efficient components pricing rule

ECPR is a regulatory rule designed to promote productive efficiency among access seekers when there is a vertically integrated access provider. It involves a two-part regulatory process. First, a price cap restrains final product prices. Given this cap, the facility owner can sell access at a price that not only recoups production costs, but also compensates for any forgone profits from final product sales due to the additional competition from access seekers. ECPR has been strongly advocated by Baumol and Sidak for electricity and local telephone access in the United States (Baumol and Sidak 1994, 1995).¹³

The logic behind ECPR is both simple and, in the correct context, compelling. As an example, consider that the marginal cost of access is \$3 per unit of the final good and the other marginal costs of providing a unit of the final good are \$5. Let the price of the final good be constrained by a price cap at \$10. A facility owner who is vertically integrated and operates as a final market monopolist would make variable profits of \$2 per unit from final goods sales. Assume that if the facility owner provides a unit of access to a competitor in the final market, then the access provider's sales in this market decline by one, for each extra unit sold by the competition. If the facility owner provides one unit of access to a competitor who uses this access to provide one unit of the final good, then the opportunity cost of providing this access is the marginal cost \$3 plus the forgone profits \$2. Under ECPR, the access charge should be set, not at SRMC, which is \$3, but at opportunity cost which is \$5.

Why is such pricing desirable? Consider a downstream competitor who is not as efficient as the incumbent integrated facility owner. For example, the competitor's

¹² In the case of access, the optimal Ramsey prices will also depend on the relative 'elasticity' of final product demand and the nature of competition between downstream firms.

¹³ Hyde and Negrin (2000) provide a recent survey on ECPR.

cost of producing the downstream product given a unit of access might be \$5.50. It would then be economically undesirable for the competitor to enter the downstream market. Such entry would result in inefficient production. However, the inefficient competitor could enter if access were priced at SRMC. If the competitor paid \$3 for a unit of access and paid an additional \$5.50 to produce the final product given the unit of access, then the total cost per unit would be \$8.50. This is \$0.50 more than the cost for the incumbent firm, but is less than the market price of \$10. The inefficient competitor can enter and make \$1.50 profit per unit. In contrast, if access prices were set by ECPR, then the total cost of production for the inefficient competitor would be \$10.50 — \$5 for a unit of access and \$5.50 to turn this unit into final product. The competitor could not make a profit at a final market price of \$10. Only if the competitor's costs of turning access into final product were at least as low as those of the incumbent facility owner, would the competitor be able to profitably enter when access is priced according to ECPR.¹⁴

ECPR can promote efficient entry and production in the final goods market. However, this is not the same as efficient production of the access service. By focusing on efficient production by access seekers, ECPR simply duplicates a function of competition. If the ultimate objective is to open final market production to competition by requiring the facility owner to provide access, then so long as all competitors (including the facility owner) can buy access at the same price, competition should drive inefficient downstream producers out of the market. If an access seeker is inefficient, then an alternative producer should be able to enter the final market, buy access and force out the inefficient producer. If the access provider is inefficient in downstream production, then they will also be forced out of final market production and will have to retreat to simply providing access to essential infrastructure services. Put simply, ECPR only does what competition would do anyway.

ECPR is a useful adjunct to downstream competition only when regulators believe that downstream competition will be either muted or distorted. For example, if regulators restrict final market entry to a single licensed firm, then there is unlikely to be effective final market competition. ECPR can be used in this situation to prevent the licensed firm entering markets where it is less efficient than the

¹⁴ The example given above is very simple. In any actual market, ECPR needs to be adjusted to retain its desirable features. For example, a unit of competitor's product need not crowd out exactly a unit of the incumbent's product (Armstrong, Doyle and Vickers 1996). The example also assumes that the final price stays constant at \$10, even if inefficient entry occurs. In contrast, if the access price was set at \$3 and the threat of inefficient entry forced the incumbent facility owner to reduce the final product price to \$8.49 in order to prevent entry, then the threat of (inefficient) entry would not lead to any actual inefficient production, but would benefit consumers by lowering the final product price (Economides and White 1995).

vertically integrated incumbent. However, a better solution to this problem may be to allow unrestricted entry downstream.

Alternatively, a vertically integrated incumbent may be able to abuse its control over access to distort downstream competition and either provide its own subsidiary with access at a more favourable price than it offers to competitors, or engage in cross-subsidisation in order to manipulate its regulatory regime (Brennan 1987). ECPR will limit these distortions. For example, it does not pay a vertically integrated producer to protect its own inefficient downstream subsidiary under ECPR, as it is fully compensated for forgone profits. However, a better regulatory solution may involve tackling the distorted competition at its source, by requiring the access producer to vertically divest its downstream subsidiary.¹⁵

A third situation where ECPR can aid efficient downstream entry is if regulated final product prices involve cross-subsidies. For example, the vertically integrated incumbent producer may be required to cross-subsidise rural consumers by charging a higher price to urban consumers. If the downstream market is opened to competition with access sold at SRMC, then new entrants will only sell to profitable market segments. New entrants will be able to undercut the incumbent in the urban market, as they do not need to cross-subsidise rural consumers. To protect the incumbent and maintain the cross-subsidy, ECPR can price access so that new competitors in the urban market compensate the incumbent for any forgone profits through the access price. However, a better regulatory solution may involve either removing the cross-subsidies from the final product prices or removing the facility owner's burden to finance these subsidies.

ECPR is most useful when political constraints prevent the introduction of fair competition in the downstream market. However, in the absence of such concerns, ECPR is equivalent to a final market price cap regime, with unregulated uniform access prices. The incentives for productive efficiency in the provision of access under a final market price cap will be similar to those that arise when access prices are directly capped. To the degree that future price caps are set according to current profits, price cap regulation will provide distorted productive incentives, similar to ROR regulation. Also, if the price caps are set too high, then there will be a loss of allocative efficiency. ECPR does not address either of these issues.

If ECPR is used without final market price caps, then the rule is the same as allowing unconstrained monopoly pricing of access. This has been recognised by the proponents of ECPR (Baumol and Sidak 1994, p. 108). However, the problem is

¹⁵ For a discussion of the arguments for and against vertical divestiture, see chapter 8 in King and Maddock (1996a). See also King (1999).

not the access pricing rule, but rather that ECPR is really a two-stage process. It is only designed to work with final market price controls in place.

The access pricing rule under ECPR is not designed to promote either allocative efficiency, efficient access production or efficient investment decisions over time. Rather, these issues are handled by final market price constraints under ECPR. The access pricing rule is designed solely to promote efficient entry in the downstream market in circumstances when other distortions exist that would impede such entry and those distortions cannot be removed for either economic or political reasons.

Long-run marginal cost pricing

Long-run marginal cost (LRMC) is the cost of increasing output by one unit when both the variable costs of production and the costs of expanding the relevant facility are taken into account. Unlike SRMC, which ignores any input costs that cannot be adjusted in the short term (eg production capacity that is already rented on an annual contract), LRMC considers the cost of production when all inputs can be varied. Because it can allow for an optimal plant capacity, LRMC will be below SRMC whenever existing production facilities are operating above their optimal capacity. However, if existing facilities are operating below optimal capacity, LRMC will exceed SRMC.

If demand for access grows steadily over time, then requiring a facility owner to price access at LRMC can induce efficient investment without any loss of allocative efficiency. Under LRMC pricing, capacity will adjust to the economically optimal level, while simultaneously setting the access price equal to the congested adjusted SRMC.

Unfortunately, if demand for access is more complicated — involving, for example, periods of peak and off-peak use a variety of different peak periods say over a daily cycle, or around a trend rate of growth — then simple LRMC pricing will promote neither allocative efficiency nor optimal investment. If there are multiple peak periods using the same access facilities, then economically optimal investment requires expanding the facility until the total value of an extra unit of capacity to all peak users is equal to LRMC. Pricing to promote efficient investment would require that the sum of all peak period prices equals long-run marginal cost. Setting price at LRMC to all users would substantially overprice access and lead to under investment. If there are both peak and off-peak users, then LRMC pricing will lead to allocative inefficiency in off-peak periods. Peak period users will have too little incentive to substitute to off-peak and off-peak users will tend to purchase too little access. Finally, if there is a single (peak period) demand that is volatile, LRMC

pricing will require non-price rationing when demand is unexpectedly high and will overprice access when demand is unexpectedly low.

LRMC pricing is most likely to be useful where there is a single, well-defined peak period for access demand. In such a circumstance, efficient investment is encouraged by requiring the access provider to sell reserved capacity during the peak period at long-run marginal cost and to expand capacity whenever demand in the peak period cannot be met at this price. SRMC pricing can be used in off-peak periods, and upfront tariffs can recover any costs associated with initial construction that are not covered by LRMC pricing.

These situations are likely to be rare in practice. In energy industries, peak use is often related to weather and cannot be predicted in advance. In water industries, new infrastructure involves lumpy investments, so that LRMC is not well defined. In urban rail systems, there are usually multiple peaks in both the morning and the evening rush hours. While LRMC tariffs may provide a useful basis for access pricing in some industries, these cases will be the exception rather than the norm.

Access pricing and asymmetric information

Access pricing inevitably involves information asymmetry between the access provider, the access seekers and the relevant regulatory authorities. Regulators will be unable to perfectly value the assets used to provide access. Short-run and long-run cost information will not be easily available and the access providers who are most likely to know relevant cost figures will often have little incentive to correctly provide this information to regulators.

Sensible access pricing procedures recognise this information asymmetry. For example, price cap regulation is based on the premise that the regulator will not be able to perfectly monitor cost reducing activities by the access provider. By allowing the access provider to retain any benefits from such activities, price cap regulation compromises allocative efficiency in order to maintain a greater degree of productive efficiency.

Economic theory provides a wide range of tools designed to overcome regulatory problems with information (Baron 1989, Laffont and Tirole 1993). The lessons from this literature are relatively simple. First, regulators need to recognise that they will often be at a disadvantage compared to access providers. There is little point requiring an access provider to set capacity according to LRMC, when the access provider has the only information about those costs for the regulator. Such a request simply invites the access provider to plead ignorance, manipulate accounting

measures of cost, and pad out LRMC measures, so that any reports will provide little indication of the true level of costs.

Second, efficient access pricing schemes will try to overcome asymmetric information and recognise information constraints that are unavoidable and minimise their effect. For example, benchmarking access against comparable facilities in Australia or overseas can provide a powerful tool to overcome information problems and improve incentives. Benchmark comparisons can be used to structure price caps or revenue caps for the access provider. Because these caps depend less on the access provider's own costs and more on the costs of related businesses, the facility owner has greater incentives to operate efficiently and less incentive to distort reported costs.

Similarly, where a number of participants in the market place may have relevant cost information, then the regulator can use competition between these participants to reveal information. Franchise auctions are a simple example of a regulatory scheme designed to elicit information from multiple providers (Demsetz 1968, Crew and Zapan 1991). Alternatively, regulators can use information provided by access seekers to control access prices. A simple example is a capacity auction for a gas pipeline, an optical fibre cable or a railway line.

Third, practical access prices will involve compromise. It is usually impractical to try and design an access pricing scheme that satisfies all the requirements of economic efficiency. A desirable access-pricing regime will tradeoff distortions, weighing each by its potential long-term detriment to welfare. Thus, a regulator may be quite satisfied with a scheme that leads to overpricing in the short term, if this loss is more than outweighed by long-term productivity improvements. Similarly, it may be desirable to allow an access provider significant profits through the upfront component of a nonlinear pricing scheme, if this leads to improved investment incentives. An important tradeoff involves administrative simplicity and regulatory complexity. While such compromises are unlikely to satisfy access providers, access seekers or final product consumers, maintaining a balance of incentives, information requirements and simplicity is most likely to provide economic benefits.

4.5 Rail access pricing

To apply the principles of declaration and access pricing developed above, consider the issue of access to railway lines. Such lines were specifically mentioned as a 'natural monopoly' in the Hilmer report (Hilmer 1993, p. 240). Rail has also been the main source of declaration decisions by the NCC.

There are at least three different types of railway lines that need to be considered for access — urban rail networks, general (long distance) freight and passenger rail, and high-volume regional rail. Consider these in turn.

Urban rail networks

Applying the tests presented in section 4.3, urban rail services are not essential services. In particular, urban rail networks fail the first test: that access (or increased access) to the service will substantially increase competition in the market for a final good or service. Urban rail transport is part of a highly competitive market for public transport services. In general, competition is so fierce that urban rail systems do not provide a standard commercial return to their government owners. Competition from other forms of public transport and from private cars means that access to an urban rail network will not substantially increase competition in the urban passenger transport market.

Urban rail systems are best thought of as ‘convenient facilities’ for the purpose of access (King and Maddock 1996b, 1999c). Urban rail probably involves a natural monopoly technology. Given the level of demand for these services, they are most efficiently supplied using a single rail network. But urban rail operates in a broader, competitive market. Reform of urban rail might involve an access regime. For example, it might be desirable to allow competing private train operators to operate on an urban rail system and an access regime would facilitate this. At the same time, alternative methods of operation, such as the use of franchise contracts, may be preferred. Regardless of the method used to operate urban rail systems, however, they are not essential facilities.

General freight and passenger rail

Like urban rail, many companies that provide general freight and passenger rail services operate in a highly competitive market, where road transport provides a viable and effective alternative. While access to general freight track services may allow some increase in competition, it is far from obvious that there will be a substantial increase in competition in general freight and passenger transport. General freight and passenger rail is best thought of as a ‘convenient facility’ for access, rather than as essential facilities.

Again, this does not mean that there is not room for reform and improvement in general rail services. Poorly run publicly-owned firms have often provided these services. There is often considerable room for improvement in the way that these services are supplied. Such improvement may involve an access regime, so private

firms can operate long distance rail services. It might be desirable to privatise the rail infrastructure. To the degree that such infrastructure involves few, if any, external costs or benefits, unregulated private ownership of the rail tracks would seem to create few economic or social concerns.¹⁶ Where rail transport does involve external effects, for example in providing transport links for people who are unable to use alternative road transport, these issues might best be dealt with through specific subsidies to the relevant passengers. Overall, general freight and passenger rail is not an essential facility and should not be subject to a general access regime and constrained by standard access pricing rules.

High volume regional rail

Some rail infrastructure involves the transport of specific bulk commodities, such as coal, wheat and iron ore. In some of these cases, there is no competitive alternative. To see whether these facilities satisfy the three essential facility tests presented in section 4.3, each test can be considered in turn.

First, would access (or increased access) to the service substantially increase competition in the market for a final good or service? In part, this depends on the relevant bulk commodity and the final markets that involve this commodity. Many commodities, such as coal and iron ore are sold into world markets, where Australia is effectively a price taker. While it is in the interest of Australian firms to produce these bulk commodities as efficiently as possible, it is not clear that allowing rail access will increase competition in any final market for a good or service. In this sense, high volume regional rail might also be a 'convenient facility'. Many bulk rail services could be reformed. Historically, State Governments have used bulk rail transport to raise revenue. Pricing of this rail transport has not been efficient, with prices considerably exceeding marginal cost. It might be better to privatise these railways, possibly selling them to a cooperative involving the firms that use the bulk transport services. Alternatively, it may be useful to create an access regime so that the rail users can operate or organise their own transport. However, once again, the rail infrastructure would not be an essential facility for the purpose of access.¹⁷

This said, there might be some circumstances where access will lead to a substantial increase in competition in a final good or service market. In these cases, the second test needs to be considered — whether the service is an input into further production and there is no alternative service or production process available to the access seeker that reasonably can be used for the production of the relevant final good or

¹⁶ King and Pitchford (1998) provide a framework for analysing privatisation decisions.

¹⁷ PC (2000, p. 162) notes that increased competition from access reforms on coal railways in regional Queensland and New South Wales has led to significant benefits.

service. Again, this test depends on case-by-case analysis. For example, the rail transport of wheat is clearly an input into further production and it is unlikely that there is any alternative process that could be economically used to transport wheat, or any other product to wheat that could be practically used as an alternative in many final goods and services.

The test will often also depend on the specific circumstances of the facility. An example is provided by Robe River's application to the NCC in 1998 for the declaration of a rail line owned by Hamersley Iron in Western Australia. This railway line was used exclusively to transport iron ore. But there were two railway lines that could be used by Robe River for the transport of their iron ore from the mine to the port — the Hamersley Iron railway and one owned by BHP. If the BHP railway was an effective alternative for Robe River, then their application to declare the Hamersley railway would fail this second test, as there would be an alternative service available that could be used for production. Robe River could negotiate access services from either Hamersley Iron or BHP. As noted above, two providers does not guarantee competition. But, if it is in fact efficient to sell access to Robe River, and the two potential suppliers do not violate the anti-collusion requirements of the Trade Practice Act, then Robe River should be able to successfully negotiate rail access off one of the two providers without the need for regulatory intervention.¹⁸

The third test — whether efficient production of the service necessarily involves an infrastructure facility with high fixed costs and relatively low operating costs so that it is likely that the development of another such infrastructure facility by any person will raise the total cost of supplying the market for the relevant final good or service — will often be satisfied for bulk rail transport.

Rail access pricing

If regulated access is used for rail services, how should it be priced? In many situations, including long distance freight and passenger rail transport and regional high-volume rail transport, the most sensible solution will involve auctioning access. PC (2000) canvasses using auctions to both price and allocate rail access.

Auctions involve little specific information for the regulator and do not require the regulator to use 'ROR' regulatory procedures. Auctions can be carried out in a

¹⁸ The case for a negotiated settlement is even stronger in this situation, as Robe River had indicated that they would build their own railway if they were unable to gain access. Hence, even if Hamersley Iron and BHP both had concerns about final market competition, denying access to Robe River would not prevent this competition.

mechanical fashion, and while a well-designed auction is not immune from collusion, it is less susceptible to collusion than many other allocation and pricing procedures. An auction process uses the information that is possessed by access seekers and forces these participants to reveal this information. For track sections where there is significant demand, an auction forces participants to reveal this demand and allows prices to reflect demand. Where there is little demand for a rail section, then this is also revealed by the auction. The track owner can then decide whether the section is economically viable over the longer term. Overall, the use of auctions to allocate and price access satisfies the key principles of good regulatory decision-making developed in section 4.4 (access pricing and asymmetric information).

The auction process requires some regulatory intervention. In particular, the regulator will need to design and oversee the auction process, and establish the parameters for the auction. For example, the regulator will need to determine the period of time between auctions. Are participants purchasing access for one year, five years or ten years? The regulator will need to set a reserve price for the auction of track sections. This reserve price should be set at SRMC including the costs of track maintenance. But the reservation price should not include any return on the value of the asset. If demand for a track section means that price exceeds SRMC but does not cover the cost of refurbishing and replacing the track in the longer term, then the owner can close down that section of track in the longer term.

The design of the auction process will clearly be critical to the successful allocation of track access. The auction should involve multiple rounds of bidding, as in standard auctions of wireless spectrum. The auction needs to allow participants to ‘build up’ complete journeys from rail sections. This means that the regulator needs to determine the relevant individual rail sections to be auctioned, both geographically and by time of day, day of the week and possibly, time of the year. It also means that the multiple rounds of the auction need to determine current highest bids by considering both the bids placed over complete journeys and bids over smaller track sections. In essence, an auction algorithm will, at any round of the auction, determine the maximum value (in terms of the consistent current bids) of all the track sections of the railway in aggregate. Such an algorithm can easily be designed for a computer.¹⁹

The regulator might also want to allow a secondary market in track sections to develop, to allow for beneficial reallocations between auctions.

¹⁹ Professor Charles Plott from CalTech illustrated the use of such a computer algorithm for rail track access in North America in a seminar at the University of Melbourne in 1999.

The use of auctions to allocate rail access has been met with some opposition. For example, it has been suggested that such an auction will be difficult to design (PC 2000, p. 181; Affleck 1999). While designing the auction will not be a simple task, it is far from obvious that the task is more onerous than attempting to apply ROR regulation to rail lines. In fact, the information required by the regulator to design a successful auction is far less than under standard regulation. Further, even if it is difficult to design an auction, the potential gains compared to traditional regulation, in terms of efficient allocation of access and pricing, are likely to more than compensate for the initial effort.

It has been claimed that auctions would disrupt existing long-term business relationships (PC 2000, p. 182). This will only be true where the value of the rail access to another firm is greater than the value to the current rail user, *even taking account of existing business relationships*. In other words, auctions only disturb existing business relationships if these relationships do not involve the most efficient use of the track. In such circumstances, it is economically desirable to ‘disrupt’ these relationships. Uncertainty over the future ability to use a certain route might also reduce the incentives for train operators to make route-specific investments. To the extent that this is an issue, it suggests that auctions should occur at longer intervals.

It has been noted that auctions may lead to very low prices for some track segments (PC 2000, p. 182). This is undoubtedly the case. And in such situations it might be desirable to close the track down in the longer term. The regulator needs to establish procedures that allow low value track to be removed from the access regime in the longer term if auction prices do not cover the on-going long-term operating costs.²⁰ Where a track is removed from the auction process in the longer term it need not be closed down. Rather, the owner of the track could still sell access to the track by individual negotiation with access purchasers. However, the price information revealed by the auction would show that demand for the track was not high enough to warrant formal regulatory intervention. The regulator might establish rules to allow such a track to be re-included in the auction at a later date, for example, if the owner requested inclusion or if there was a significant change in circumstances that would lead to a significant increase in demand.

A successful auction that included long distance freight and passenger transport would require coordination across State boundaries. Because of this, an auction could not be done on a state-by-state basis, but would need to be carried out on a national basis.

²⁰ Formally, a track may be removed from the auction process if the total revenues accruing to the track over, say, one year, do not exceed the avoidable costs of operating that track over the same period.

If an auction was not used to price and allocate rail access, then standard regulatory access pricing would be needed. Such pricing should be based on SRMC with, say, a per rail segment annual fixed charge to cover fixed costs. If there were multiple access seekers for any rail segments, then the regulator would need to develop rules to allocate access.

New rail developments

There are a number of new rail developments that have been proposed involving both freight and passenger transport. Examples include the Alice Springs to Darwin railway and the Sydney to Canberra fast rail project. These projects involve substantial degrees of uncertainty. It is likely that the projects are either only marginally viable on a commercial basis or require government funding to make them feasible. Following the discussion in section 4.3 (undertakings and new investment), it would be sensible for these projects to be exempt from any access regime.

It could be argued that, for these major projects, regulated infrastructure access is desirable *ex post*. Even ignoring the potential for an access regime to deter investment, the economic arguments to include new rail infrastructure in an access regime appear weak. Bottleneck infrastructure causes a problem when it involves an essential facility. But a new railway that does not presently exist cannot, by definition, be an essential input for any existing downstream producer. These producers have been operating without any rail services and to the degree that rail investment creates new competition to supply inputs, the relevant downstream firms can only be made better off.

New producers, for whom the railway is a bottleneck facility, may develop after the railway is built. But again this is usually not a problem requiring regulatory intervention. Because new producers develop their facilities after the railway, there is unlikely to be an essential facility issue. The new producers will negotiate with the railway to gain long-term transport contracts before they commence any investment. The new producers will be in a strong position to bargain with the track owner — after all, the track owner has sunk their investment, while the rail service purchaser has not yet made its investment.

4.6 Conclusion

This paper has considered a variety of issues in essential facility access, including the structure of regulations and pricing. Overall, there seems to be significant room for improvement in Australian access regimes. The general regime in Part IIIA of

the Trade Practices Act does not properly capture essential facilities. Alternative tests are proposed above. Access pricing has often followed ROR regulation in Australia, but this will often be inefficient. When considering rail transport, access may be part of general reform, but often it is not essential facility access.

References

- Affleck, F. 1999, 'Effective competition on rail: practical issues', Chapter 11 in Productivity Commission, *1999 Industry Economics Conference: Regulation, Competition and Industry Structure*, Conference Proceedings, 12–13 July, AusInfo, Melbourne, pp. 181–98.
- Armstrong, M., Doyle, C. and Vickers, J. 1996, 'The access pricing problem: a synthesis', *Journal of Industrial Economics*, vol. 44, no. 2, pp. 131–50.
- Averch, H. and Johnson, L. 1962, 'Behavior of the firm under regulatory constraint', *American Economic Review*, vol. 52, pp. 1052–69.
- Baron, D. 1989, 'Design of regulatory mechanisms and institutions', Chapter 24 in Schmalensee, R. and Willig, R. (eds), *Handbook of Industrial Organization*, vol. 2, North-Holland, Amsterdam.
- Baumol, W. and Sidak, J.G. 1994, *Towards Competition in Local Telephony*, MIT Press, Cambridge, Massachusetts.
- 1995, 'Stranded costs', *Harvard Journal of Law and Public Policy*, Summer, pp. 837–49.
- Brennan, T. 1987, 'Why regulated firms should be kept out of unregulated markets: understanding the divestiture in United States v AT&T', *Antitrust Bulletin*, vol. 32, no. 3, pp. 741–93.
- Carlton, D. and Perloff, J. 1994, *Modern Industrial Organization*, Harper Collins, New York.
- Crew, M. and Zapan, M. 1991, 'Franchise bidding for public utilities revisited', Chapter 10 in Crew, M. (ed.), *Competition and the Regulation of Utilities*, Kluwer, Amsterdam.
- Demsetz, H. 1968, 'Why regulate public utilities?', *Journal of Law and Economics*, vol. 11, pp. 55–65.
- Doyle, C. and Maher, M. 1992, 'Common carriage and the pricing of electricity transmission', *Energy Journal*, vol. 13, no. 3, pp. 63–94.
- Economides, N. 1997, *The tragic inefficiency of the M-ECPR*, New York University, unpublished.

-
- and White, L. 1995, ‘Access and interconnection pricing: how efficient is the efficient components pricing rule?’, *Antitrust Bulletin*, vol. 40, no. 3, pp. 557–79.
- Freebairn, J. and Trace, K. 1992, ‘Efficient railway freight rates: Australian coal’, *Economic Analysis and Policy*, vol. 22, pp. 23–38.
- Gans, J. 1998, ‘Regulating private infrastructure investment: optimal pricing of access to essential facilities’, Working Paper no. 98-13, Melbourne Business School, University of Melbourne.
- and Williams, P. 1999, ‘Access regulation and the timing of infrastructure investment’, *Economic Record*, vol. 79, no. 229, pp. 127–38.
- Hilmer, F., Raynor, M. and Taperell, G. (The Independent Committee of Inquiry) 1993, *National Competition Policy*, AGPS, Canberra.
- Hyde, C. and Negrin, J.L. 2000, ‘Access pricing: a survey’, Research Paper no. 735, Department of Economics, University of Melbourne.
- Kahn, A. 1988, *The Economics of Regulation*, MIT Press, Cambridge, Massachusetts.
- King, S. 1997a, ‘National competition policy’, *Economic Record*, vol. 73, no. 222, pp. 270–84.
- 1997b, ‘Access pricing under rate-of-return regulation’, *Australian Economic Review*, vol. 30, no. 3, pp. 243–55.
- 1999, ‘Price discrimination, separation and access: protecting competition or protecting consumers?’, *Australian Journal of Management*, vol. 24, pp. 21–35.
- and Maddock, R. 1996a, *Unlocking the Infrastructure: The Reform of Public Utilities in Australia*, Allen and Unwin, Sydney.
- and — 1996b, ‘Competition and almost essential facilities: making the right policy choices’, *Economic Papers*, vol. 15, no. 3, pp. 28–37.
- and — 1999a, ‘Issues in access’, Chapter 2 in Productivity Commission, *1999 Industry Economics Conference: Regulation, Competition and Industry Structure*, Conference Proceedings, 12–13 July, AusInfo, Melbourne, pp. 19–30.
- and — 1999b, ‘Light handed regulation of access in Australia: negotiation with arbitration’, *Information Economics and Policy*, vol. 11, pp. 1–22.
- and — 1999c, ‘Rail access and rail reform’, Chapter 16 in Productivity Commission, *1999 Industry Economics Conference: Regulation, Competition and Industry Structure*, Conference Proceedings, 12–13 July, AusInfo, Melbourne, pp. 219–29.

-
- and Pitchford, R. 1998, 'Privatisation in Australia: understanding the incentives in public and private firms', *Australian Economic Review*, vol. 31, no. 4, pp. 313–28.
- Laffont, J. and Tirole, J. 1993, *A Theory of Procurement and Regulation*, MIT Press, Cambridge, Massachusetts.
- Mankiw, N.G. and Whinston, M. 1986, 'Free entry and social inefficiency', *Rand Journal of Economics*, vol. 17, no. 1, pp. 48–58.
- McAfee, R.P. and Schwartz, M. 1994, 'Opportunism in multilateral vertical contracting: nondiscrimination, exclusivity and uniformity', *American Economic Review*, vol. 84, no. 1, pp. 210–30.
- Mitchell, B. and Vogelsang, I. 1991, *Telecommunications Pricing: Theory and Practice*, Cambridge University Press, Cambridge.
- NCC (National Competition Council) 1996, *The National Access Regime: A Draft Guide to Part IIIA of the Trade Practices Act*, NCC, Melbourne, August.
- 1997, *Application for Declaration of Central Rail Freight Services, Brisbane to Cairns Rail Corridor: Reasons for Decision*, NCC, Melbourne, 3 June.
- Panzar, J. 1989, 'Technological determinants of firm and industry structure', in Schmalensee, R. and Willig, R. (eds), *Handbook of Industrial Organization*, vol. 1, North-Holland, Amsterdam.
- PC (Productivity Commission) 2000, *Progress in Rail Reform*, Report no. 6, AusInfo, Canberra.
- Rey, P. and Tirole, J. 1996, A primer on foreclosure, IDEI, France, unpublished.
- Slater, M. 1989, 'The rationale of marginal cost pricing', in Helm, D., Kay, J. and Thompson, D. (eds), *The Market for Energy*, Clarendon, Oxford.
- Waterson, M. 1987, 'Recent developments in the theory of natural monopoly', *Journal of Economic Surveys*, vol. 1, no. 1, pp. 59–80.

Discussant — *Ed Willett*

National Competition Council

D4.1 Scope of regulation

It seems to be common ground that the regulation of essential facilities should be about providing for the shared use of natural monopoly infrastructure services in order to promote competition in markets that depend on that infrastructure. There is also unlikely to be much difference between us on the sorts of infrastructure services that it is desirable to make subject to access rules. However, the paper seems to argue that are, in many cases, these natural monopoly infrastructure services would not accord with the paper's definition of 'essential facilities'. The consequence of that seems to be that the paper would then suggest that regulation should not occur.

The paper argues that regulation is only appropriate to promote competition in final product markets, however defined, rather than markets for inputs to the production of final products. I do not agree with this approach because it bundles efficiency and competition issues. Increased competition in input markets via access regulation of natural monopoly infrastructure will provide benefits to the community by reducing the price and/or improving the quality of inputs. Those benefits are likely to flow through to final product markets through efficiency gains delivered to, or derived in, those final markets. But it is a very different thing to suggest that the benefits of access regulation should only be tested according to whether it enhances competition in a final product market.

D4.2 Uneconomic to develop another facility

The paper points out that the NCC has applied a 'social' test to the interpretation of the term 'uneconomic to develop another facility to provide the service' in the criteria for a declaration recommendation.²¹ The paper also suggests, correctly, that by applying this approach, the NCC is regarding the relevant criterion for declaration — criterion (b) of s. 44G(2) of the Trade Practices Act — as an explicit test of natural monopoly.

²¹ This approach to the Council's application of this criterion became prominent in the context of the Robe River declaration application for certain rail services in the Pilbara. Prior to this, the Council had indicated that the declaration criteria were designed, among other things, to identify infrastructure with natural monopoly characteristics. However, in earlier matters, the distinction between the private and social tests of 'uneconomical' was not critical, and evidence of private viability was often also evidence of social viability.

The paper goes on to suggest that the current wording of the Act may not allow for the NCC's interpretation. However, the paper omits the fact that the Australian Competition Tribunal, in the Sydney Airports matter, clearly endorsed the application of a social test and expressly considered a social test by reference to the notion of natural monopolies.

Whilst the Tribunal is not a court, all questions of law (including questions of statutory interpretation of provisions of the Trade Practices Act) are determined by the Federal Court Judge presiding in the matter. Further, whilst there are review rights to the Federal Court and potentially further, the Tribunal has been very influential in judicial consideration of the Trade Practices Act since its inception in 1974.

In view of the present state of the authorities, the better view is that criterion (b) is a test of natural monopoly expressed as a social test of uneconomic to develop another facility.

D4.3 Promotion of competition

Of course, the uneconomic to develop another facility test is just one of six criteria for declaration. The existence of natural monopoly is not sufficient cause for a declaration recommendation. In particular, for a declaration recommendation to be made, criterion (a) requires that the Council be satisfied that access (or increased access) to the relevant natural monopoly service would promote competition in another market.

D4.4 Proposed reformulation of criteria

This brings me to the paper's proposed redrafting of criteria (a) and (b) into three new criteria: these are described in the paper as criteria (a), (b) and (c).

I will start with the paper's proposed criterion (c), which is the purported new natural monopoly test. As the paper points out, drafting legislation to deal with a concept such as natural monopoly is a difficult exercise and I cannot predict how courts are likely to view the words suggested. I do not have any difficulty with the particular words used, except for the focus on 'final product' markets, which I will discuss a little later. But if what we want is a natural monopoly test, I do question whether it is sensible to replace the current one-line test, which has already been endorsed by the Tribunal as a natural monopoly test, with a four line test which has no authority to support it and, prima facie, a change to the legislative wording

would be evidence of a desire to change from the accepted interpretation of the existing wording.

Turning now to the proposed new criterion (a). This proposes two significant changes to the current promotion of competition test. I do not think either of these changes would improve the criteria for declaration.

The first proposes a change to focus the competition test on final product markets. Clearly, competition in final product markets creates discipline throughout the production chain. For example, progressive tariff reductions and internationalisation in the traded goods sector in Australia has led to improvements in the performance of labour markets. But I think that it is going too far to suggest that competition in final product markets renders competition in, and the application of competition policy to, input markets irrelevant.

The paper recognises that in the Hunter Valley, successive NSW governments have extracted substantial rents from coal mining companies through excessive pricing of rail services and that access regulation has substantially improved outcomes. The coal from the Hunter Valley is either exported or used in the generation of electricity in the region. According to the redrafted competition test, it would be necessary to demonstrate a ‘substantial increase’ in competition in either:

- the National Electricity Market, where there are many electricity generators who do not rely on coal haulage services in the Hunter Valley; or
- some overseas final product market that relies on the market for coal, again where there are many producers who do not rely on coal haulage services in the Hunter Valley.

Thus, the use of the proposed competition test would suggest that declaration (or any access regulation) of rail track services for coal haulage in the Hunter Valley would not be available. This is despite the fact that, clearly, competitive disciplines in related final product markets have not sufficiently disciplined the provision of rail track services in the Hunter Valley to avoid welfare reducing rent-seeking behaviour by the monopolist (regardless of whether these rents are extracted through high profits, cost padding or an ‘easy life’).

Similarly, the paper notes that access regulation applied to rail infrastructure can improve outcomes in relation to the other bulk commodity freight (such as wheat) and long distance general freight services.²² Again, these rail track services may not

²² The paper suggests that its proposed tests for declaration may be satisfied in relation to the rail haulage of wheat. But its not clear to me that this is the case: surely, final product markets using wheat as an input are likely to be quite broad and diffuse, and unlikely to rely on access to a specific set of rail infrastructure.

satisfy the paper's proposed tests for declaration, but would meet (and have in some instances have met) the current test for declaration as applied by the NCC and the Tribunal.

The second proposed change is to replace the words 'would promote competition' with the words 'will substantially increase competition' with a view, I think, to requiring a higher competition enhancement hurdle for declaration. I am not sure the new test would achieve this objective. In the Sydney Airports decision, the Tribunal explicitly distinguished the notions of promoting competition and increasing competition. The Tribunal considered that, in the context of the declaration criteria, promoting competition 'involves the idea of creating the conditions or environment for improving competition from what it would be otherwise'.²³

Essentially, access regulation is about dealing with structural impediments to competition in markets that rely on services provided by natural monopoly infrastructure. In some cases, the natural monopoly service will be, in effect, the exclusive source of competition in the other market, because no other sources of competition are feasible. This was the situation in the Sydney Airports matter, where the Tribunal considered that removing the structural impediment to entry in international freight handling services in Sydney by providing access to Sydney Airport promoted competition in those freight handling services markets.

In other cases, there may be other sources of competition in the related market. In these cases, consideration of the promotion of competition test by the Council has been more difficult. Examples have arisen in relation to:

- interstate rail services (and whether the availability of road haulage services would mean that access to rail track would not promote competition in relevant interstate freight forwarding services markets); and
- the application for coverage of the Eastern Gas Pipeline under the National Gas Code (and whether access to two pipelines which are natural monopolies within their respective gas transmission services markets would promote competition in the south-east Australian gas sales market).

In these cases, it is possible that the focus on structural impediments to competition reflected in the Tribunal's application of the promotion of competition test provides a higher competition enhancement hurdle than a substantial increase in competition test.

²³ Australian Competition Tribunal (2000 ATPR 41-754 at 40,775).

In any case, the paper proposes to exclude the possibility of declaration where there is more than one source of competition in the final product market by the inclusion of another new criterion. Notably, the paper recognises that access regulation can, and often does, improve outcomes where two natural monopolies provide input services to a final product market, but considers that declaration should be ruled out in these circumstances because it appears to be a neat, if imperfect, place to ‘draw the line’. I suggest that a better place to ‘draw the line’ is closer to where it is drawn at present: the regulation of natural monopoly infrastructure to promote competition in another market.

D4.5 Market definition

There is another element to this focus on final product markets that I would like to touch on. The focus seems to reflect a view that not only does competition in input markets not matter, but there is some doubt about whether such input markets exist as separate markets at all. This goes to the delineation of functional markets in the supply chain, and is fundamentally important to the regulation of natural monopolies. The view that functional markets do not exist, or that competition in functional markets is not important, was reflected in the expert evidence of Professors Parry and Maddock in the Sydney Airports matter. The view is not consistent with trade practices law in Australia, and was soundly rejected by the Tribunal in favour of the views expressed by Professors Williams and Ergas in that matter on the role and significance of functional markets. Notably, in its decision, the Tribunal found that access via declaration would promote competition in the functional input markets relating to the handling of international air freight.

Further, on the question of functional market delineation, the proposed redrafting of the competition test in the paper may have diluted the need to demonstrate that competition is promoted in a market separate from the market for the natural monopoly service. I think this would be a serious backward step.

Access regulation is about, in large part, distinguishing between competitive infrastructure (including potentially competitive infrastructure) and natural monopoly infrastructure. Access to natural monopoly services is regulated to promote competition in distinct dependent areas of economic activity. Access regulation tends (and indeed is intended) to entrench the existing structure within a natural monopoly market. This reflects the fact that, in relation to natural monopolies, the development of competing infrastructure is undesirable. The separate market test ensures that access regulation is not applied to infrastructure service markets that are competitive, or potentially competitive, in order to promote competition in those markets. Access regulation applied to infrastructure, even for a

limited time, in the hope that such regulation will itself promote the development of competing infrastructure does not make any sense.

It is notable that in telecommunications access regulation under Part XIC of the Trade Practices Act, there is no separate market or equivalent test. This is despite the fact that, of the infrastructure that is currently subject to access regulation, it is in telecommunications that the development of competitive infrastructure would appear most likely. The danger is that without a separate market test or some equivalent, access regulation would be applied to potentially competitive infrastructure in telecommunications because current levels of competition in the market for services provided by that infrastructure are viewed as deficient. By deterring investment in competing infrastructure, access regulation in these circumstances is likely to be counterproductive.

D4.6 Other comments

Some quick comments on other points in the paper.

- The paper suggests that Part IIIA cannot deal with situations where access is currently provided, but at high prices. I do not agree. In fact, several declaration applications have been submitted to secure 'increased access' by people who currently have access, but on unfavourable terms.
- The paper refers to the risk of negotiated access leading to a sharing of rents by infrastructure owner and access seeker. There is a theoretical possibility of this, although I believe that the fact that declaration provides access for all comers substantially reduces, rather than enhances, this risk. Further, in five years experience with Part IIIA, I cannot think of any example where there was even a remote prospect of such rent-sharing. Requiring regulatory approval for all access contracts to deal with such a remote risk appears heavy-handed.
- The paper discusses the problem of ex ante and ex post risk assessment and the asymmetric impact of access regulation. This is an issue for any access regulation and is currently the subject of some thought. It may be that, in some circumstances, the costs of access regulation may exceed the benefits even in the case of natural monopoly infrastructure where access would promote competition in another market. The public interest test in the criteria for declaration should ensure the right result in such circumstances. There is also the opportunity, as the paper explains, for someone planning risky green-fields investment to submit an undertaking to the ACCC prior to committing to the investment. The third option, and the approach adopted in the case of the Tarcoola/Darwin railway, is for the relevant government/s to enact a specific regime for a new project, and seek certification by the Council, if necessary, to

protect the project from declaration. Problems associated with ex post risk assessment and the asymmetric impact of access regulation were prominent in the Tarcoola/Darwin railway project. Nonetheless, they were resolved during the certification process in ways that provided a robust set of access rules.

D4.7 Cultural change

I will conclude with saying something about access regulation of airports, particularly in light of the Sydney Airports decision. Often I think that access regulation is about changing the culture of doing business, particularly in relation to vertically separated, publicly-owned infrastructure.²⁴ During the course of the Sydney Airports matter, I found it interesting to observe the gradually changing attitudes. By the time of the Tribunal hearing, the owner of Sydney airport, the Sydney Airport Corporation Ltd (SACL), recognised many of the problems associated with limiting access to airport services. Indications are that the decision by the Tribunal has further shifted views in SACL toward a culture of doing business to maximise utilisation of Sydney Airport.

In this context, it is interesting to note that no further issues about access to Sydney Airport have arisen formally, or even informally as far as I know. This is despite the fact that two new domestic airlines have either started operations, or taken steps toward starting operations. Both appear to have successfully negotiated access to Sydney Airport.

There may be a number of reasons why these negotiations have been concluded without formal access regulation, other than the threat of declaration. That of itself does not mean that more formal access regulation is not desirable. But the prospect that light-handed regulation through the threat of declaration alone may be sufficient regulation does warrant some consideration.

²⁴ The notion of vertical separation can be problematic. In respect to access regulation, the important consideration is whether the natural monopoly infrastructure owner has (or may develop) any interests (equity, contractual or otherwise) in dependent markets.

General discussion

The discussion focused on:

- competition in input markets;
- the role of the Australian Competition Tribunal; and
- Part IIIA access regime in the context of rail.

Competition in input markets

Stephen King clarified that he did not believe competition in input markets should be ignored, but rather, that if competition increased in an input market, the benefits would only exist if they fed through to the final market. Increasing competition at one level would have no effect in a three-stage vertical chain, if all the benefits were absorbed at another level before reaching customers. Thus, a focus on final markets was more useful.

The role of the Australian Competition Tribunal

It was noted that the discussant placed great emphasis on the decisions of the Australian Competition Tribunal. But tribunals, like courts of law, can change their personnel and their opinions. There was agreement among a number of participants that the Tribunal had made sensible decisions with regard to the ‘uneconomic development of a facility’ test, but that the Tribunal’s current stance may indeed change at some point in the future. Ed Willett suggested that the Trade Practices Act should be amended when it was not being interpreted in the intended way.

Stephen King stated a preference for incorporating a simple natural monopoly test into the Trade Practices Act, rather than relying on sensible decisions of the Tribunal. An appropriate test would look for big fixed costs, low operating costs and little competition anywhere, rather than simply asking if the market was a natural monopoly.

Part IIIA access regime in the context of rail

It was suggested that there may be several situations where access was a useful reform measure, but this should not necessarily be dealt with under Part IIIA of the Trade Practices Act. In the example of Hunter Valley Rail, Stephen King acknowledged that there was need for reform. The State Government which operated Hunter Valley Rail set prices at an inappropriate level. A profit-

maximising monopolist would have done a much better job of getting coal to the port than did the State Government. However, a need for reform did not necessarily mean that access provisions should be a part of this reform, nor that it should be dealt with by Part IIIA. Stephen King referred to King and Maddock (1996b) that showed how alternative methods of reform might better deal with ‘monopoly’ problems in vertical production chains.

A participant referred to the example of FreightCorp. Although being Government-owned, FreightCorp must seek access from the track owners in New South Wales and from vertically integrated entities elsewhere in Australia to create a national market for bulk commodities. It was suggested that failure in this market was not a result of Part IIIA or the declaration process, but rather insufficient, or inappropriate, microeconomic reform. The States had done whatever was necessary to establish their own access regimes, but had failed to work to a common benchmark to create a national arrangement.

Reference

King, S. and Maddock, R. 1996b, ‘Competition and almost essential facilities: making the right policy choices’, *Economic Papers*, vol. 15, no. 3, pp. 28–37.

5 Regulating access to telecommunications networks

Robert Albon[†]

Australian National University

5.1 Introduction

A feature of the traditional organisation of utilities is complete statutory monopoly protecting the utility from all competition. However, this has presented a substantial dilemma for policymakers interested in economic efficiency. On the one hand, the objective of excluding competition is to prevent wasteful entry in natural monopoly circumstances, where duplication could result in higher overall costs. On the other hand, banning competition is not always the right overall approach. In particular, not all areas of a utility are naturally monopolistic, and even in areas where natural monopoly is present, competition may — on balance — be beneficial. Further, competition can have various vivifying effects, resulting in net benefits, particularly by placing downward pressure on costs, acting as a spur to innovation and facilitating the creation of benchmarks for performance appraisal. Competition can also undermine a utility's ability to practice cross-subsidisation, therefore bringing about a more efficient pricing structure. All of these arguments for and against competition have been made in relation to Australian telecommunications.

Competition for a utility is seldom introduced by the complete removal of statutory monopoly in one fell swoop and without associated regulatory and organisational changes. The utility needs to be viewed as a grouping of inter-related production components, some of which are naturally monopolistic and others that are 'naturally competitive'. For efficiency, the different areas need to be subject to different treatment. In areas that are not naturally monopolistic, competition can be introduced directly, simply by removing statutory barriers to entry for some or all of

[†] At the time of writing, Robert Albon was Senior Lecturer in Economics and Director of the Centre for Applied Economics at the Australian National University. Robert Albon is now Senior Economic Adviser (Regulatory) with the Australian Competition and Consumer Commission. Views and positions taken are his own and do not necessarily coincide with those of the ACCC.

the utility's markets and therefore allowing entrants to compete. However, where natural monopoly is present, there are likely to be 'bottlenecks' in production (or 'essential facilities') still controlled by an incumbent. Rivals will have to gain access to these monopoly areas for effective competition to occur. To enable this, the incumbent may have to be forced to provide their rivals with access to the monopoly elements on 'favourable' terms.

The first major utility area in Australia subject to comprehensive reform was the telecommunications industry. Forcing the incumbent to provide access to its rival(s) at regulated rates facilitated network competition. The beginning of this experience preceded a more general approach to utility competition — National Competition Policy (NCP) — flowing from the 1993 Committee of Inquiry into National Competition Policy (Hilmer 1993).¹ This too was an access-based approach to competition. This is now enshrined as Part IIIA of the *Trade Practices Act 1974* (Cwlth), while the telecommunications provisions (which are slightly different) are in Part XIC.

This paper considers the following aspects of the access regime in Australian telecommunications:

- the antecedents of the access pricing regime now in operation;
- a description of the arrangements for facilitating access to networks of carriers and the arrangements for determining the price of a declared service that came into operation in 1997;
- an evaluation of the efficiency of the existing arrangements in the light of established principles of efficient access pricing and structural considerations; and
- the impacts of new technologies ('convergence') on the efficiency of the access regime and, in particular, whether new technologies will erode the need for access regulation.

5.2 A brief history of telecommunications reforms

Until 1975, internal telecommunications services were provided by an operating unit (the 'Australian Post Office' or APO) within the Postmaster-General's (PMG) Department that had a complete statutory monopoly precluding virtually all activity by any private or government-owned bodies. The APO also operated postal services. International telecommunications services were exclusive to the

¹ For a description and appraisal of this process, see King (1995, 1997).

independent Overseas Telecommunications Commission (OTC), established in 1946.

The original 'Telecom' (initially the Australian Telecommunications Commission) resulted from the separation and removal from departmental operation of the telecommunications and postal parts of the operational section of the PMG in 1975. These changes followed recommendations of the 1974 (Vernon) Commission of Enquiry into the Australian Post Office. Telecom remained under government ownership and continued to have a statutory monopoly on all aspects of providing telecommunications services, including customer premises equipment (CPEs) and value-added network services (VANS). While subject to government pricing regulation and other constraints on its operations, Telecom itself acted as the 'technical' regulator of the industry.

Domestic reticulation of OTC's international traffic was handled by the APO and later by Telecom. The Australian domestic communications satellite system (AUSSAT) was formed in 1981 to own and operate Australia's communications. AUSSAT's functions included satellite transmission of broadcasting services and the provision of private network services involving interconnection with Telecom's network for the origination and termination of calls.

Corporatisation of Telecom in the late 1980s gave it a more commercial focus, with greater operational freedom, clearer objectives and a more business-like structure. However, it did not involve structural division into arms-length business units. The 1988 Ministerial Statement resulted in the opening up of most CPE and VANS areas to competition. Responsibility for regulation was removed from Telecom with the formation of an independent industry-specific regulator, the Australian Telecommunications Authority (AUSTEL).

AUSTEL commenced pricing regulation based on a 'CPI - X' price cap (restricting the increase in the weighted-average of a basket of telecommunications services to rise no more than the CPI less a specified percentage) with individual sub-caps on some prices. A requirement for uniform geographic pricing remained, meaning, in particular, that subscriber trunk dialling (STD) prices were based on distance and time of day, without any allowance for density or thickness of traffic on the route.

Another review (Review of Structural Arrangements, or ROSA) occurred during 1989-90. ROSA was concerned with the relationships between the three carriers at the time — Telecom, OTC and AUSSAT. ROSA, combined with other considerations, culminated in another Ministerial Statement in 1990, by the Minister for Transport and Communications, Kim Beazley. The principal decisions made around this time were as follows.

-
- Telecom was amalgamated with OTC forming 'Telstra'. An International Business Unit continued OTC's role within Telstra.
 - AUSSAT was fully privatised, with a complete carrier licence attached to the sale.
 - The issuing of this licence allowed network competition between Telstra and the second licensed carrier, with a guarantee of a duopoly (no new competition) until 1997. The purchaser of AUSSAT, Optus Communications, was guaranteed regulated access to Telstra's local network to reticulate its calls, with AUSTEL as the regulatory body. Optus was also licensed to operate a second mobile telecommunications service.
 - A third mobile licence was allocated to Vodafone, which chose only to operate a digital mobile service. (Telstra and Optus offered both digital and analog services.)
 - Licences were granted to various service providers (SPs) offering VANS and re-selling services purchased from Telstra.

Telstra's structure has changed considerably over the years. It has progressively been subjected to more competition, been corporatised, and assumed more functions, partly through technological change and partly from the amalgamation with OTC. It has moved from what was still a state-based management structure to one based more on functions, and this evolution is continuing. Telstra still retains a largely 'integrated' organisational structure.

The divisions and sub-divisions do not operate separately as independent profit centres. Transactions conducted between them are not at 'arm's length', there is no apparent internal pricing structure, and all are ultimately responsible to the one Board of Directors.

There were four important changes to the telecommunications arrangements commencing in 1997.

- First, the network duopoly ended and new access arrangements commenced. (These form the main focus of this paper and are considered in detail later.)
- Second, telecommunications was no longer exempt from Part IV of the Trade Practices Act. The *Telecommunications Act 1991* (Cwlth) exempted telecommunications from the Trade Practices Act and imposed telecommunications-specific competition rules. The new Part XIB of the Trade Practices Act continues specific competition rules for telecommunications. This Part is aimed at allowing the Australian Competition and Consumer Commission (ACCC) to act more quickly against anticompetitive conduct than it could under Part IV. In contrast to the old legislation, there are no restrictions on 'normal

competitive conduct’, introducing ambiguity into the operation of the Trade Practices Act. Competition notices issued by the ACCC that notify contravention of ‘the competition rule’ are ‘effects-based’, not reliant on a ‘purpose test’. The specific prohibition against pricing discrimination by Telstra that was contained in the Telecommunications Act ceased and was replaced by the general limitations on anticompetitive behaviour of the ‘competition rule’ outlined above.

- Third, existing ‘CPI – X’ price capping regulation continued. Compliance with the regulation is the responsibility of the ACCC. ‘X’, which is set to reflect total factor productivity growth, has been progressively increased over the 10 years of the regulation. Some productivity growth appears to have flowed through to excessively higher profits and excessive costs rather than price reductions. Sub-caps and explicit regulations have prevented substantial price restructuring. However, the latest round of price capping has placed customer access and local call prices in a sub-basket, allowing some efficiency-enhancing adjustment of prices. Telstra must offer an untimed charge for local calls and the Minister for Communications retains direct powers over variations in prices not covered by the price capping regulations.
- Fourth, one-third of Telstra was sold in a public float in late 1997. The Government sold another tranche of Telstra shares in 1999, taking the privately-owned portion to just under 50 per cent.

5.3 Antecedents of the current access arrangements

OTC-Telecom

Substantial information about the access pricing arrangement was publicly available until 1988-89, including in OTC Annual Reports, Alston (1988), the IAC (1989), the PSA (1989) and the BTCE (1993). After 1988-89, OTC supplied less information about how much it paid Telecom for access to its network, and, since the amalgamation of Telecom and OTC, no quantitative or qualitative information on the internal pricing arrangements is publicly available.

According to the BTCE (1993, p. 51), Telecom charged OTC a ‘flat rate of about 20 cents per minute for the domestic component of the international calls irrespective of the domestic origin or destination of the call, or time of day’. Similarly, the IAC (1989, pp. 149–50) stated that, ‘based on evidence from Telecom, it appears that this charge is a flat 20 cents per minute, irrespective of the domestic origin or destination of the international call, or time of day’. The PSA (1989, p. 12)

indicated that the 'terminal charge is differentiated according to traffic direction (incoming or outgoing) and type of connection (automatic or operator-assisted)'.

It now appears that the average rate paid to Telecom may have been higher than 20 cents, perhaps around 30 cents per minute. OTC's 1988-89 *Annual Report* states that it made a \$236.3 million payment 'to Telecom for domestic transmission' (p. 40). On the basis of OTC's both-way total of 785 086 telephone and Telex minutes, this works out at 30.01 cents per minute. Similarly, the PSA (1989, p. 33) calculates an average of 30 cents per minute, but notes that some of OTC's payments to Telecom were lump-sum rather than per-minute.

While Telecom's costs of reticulating calls varied substantially by distance, thickness of route and time of day, the rate charged to OTC was apparently in excess of the cost of even the most costly reticulations, and certainly much in excess of the cost of local reticulations. As Telecom did not operate an international call service, this excess of the interconnect price over the cost of provision cannot be explained in terms of a payment for lost contribution to overheads by Telecom. The most likely explanation is that the payment represented a form of 'profit sharing' by the two carriers.

The access regime had implications for both consumer welfare and economic efficiency. If Telecom had applied a cost-based pricing structure, and if this had been passed on by OTC in lower prices to end users, consumers would have gained and economic efficiency would have been enhanced. (A high marginal efficiency gain would flow from the large wedge between price and relevant marginal cost interacting with a highly elastic demand.)

Reticulation of traffic from AUSSAT and leased lines

For private network traffic carried long-distance by AUSSAT or on a leased line, the interconnect policy developed by Telecom involved a two-part pricing structure and was designed to prevent third-party use of the private network. Private networks were allowed to interconnect at both ends, but traffic was not allowed to go out of and back into Telecom's public network. Traffic entering the Telecom network was charged on a per-call basis at the price of an (untimed) local call. In addition, each connected line was charged an annual fee of between \$600 and \$3 300, depending on distance. The proportionate scale of charges exactly followed the STD distance-related pricing structure. The structure was later changed to a flat fee of \$2 200 per interconnected line, with intrusions into the local network still charged at the local call rate.

The rationale for this pricing scheme was that the fixed cost covered the loss of contribution to CSOs by the diversion of STD traffic into private networks, and that the charge per call reflected the cost of each intrusion into the public network. In the initial version, the fixed pricing structure discriminated against longer-distance private network traffic. The variable component was above per unit cost and did not reflect cost differences between peak and off-peak use of the local exchange network (LEN).²

Initial Optus access arrangements

When Optus commenced its plans for entry in 1991, the two carriers could not agree on the conditions of access, and AUSTEL was instructed to report to the Minister on the conditions under which Optus would be allowed to interconnect. The Minister determined that the interconnect agreement was to be based on 'directly attributable incremental cost' (DAIC) (box 5.1).

AUSTEL received advice from a variety of sources. Telecom argued for a wholesale price of about 6 cents per minute at each end of the call. The main telecommunications union argued for a similar price. Optus suggested a substantially lower price.

AUSTEL's recommendations, released in a three volume report in July 1991, were for an average price somewhere between these extremes, at 3.14 cents per minute per end. There was some slight allowance for differences in costs by region and according to peak and off-peak. The price included an estimate of 'long-run marginal cost' plus a contribution (0.52 cents per minute) to the cost of the CAN.

King and Maddock (1996a, p. 144) contend that the determined price reflected long-run variable costs. While the method AUSTEL used for isolating the costs attributable to the entrant is not clear, in effect, the estimate of long-run marginal cost (averaging 2.64 cents per minute) is roughly equal to average total cost of the LEN. The estimate of average cost across *all* minutes for the LEN in 1995-96 in Albon, Hardin and Dee (1997, Appendix A) is 2.5 cents per minute. In 1991 values, the total cost of the LEN would have been less, but the total number of minutes was also substantially less. On balance, the average cost was probably about the same.

² The LEN is the set of equipment for switching and inter-exchange carriage of telecommunications traffic. It carries local call traffic, as well as reticulating higher level traffic (eg long-distance calls) into and out of the customer access network (CAN). The LEN and the CAN together make up the 'local network'.

Box 5.1 Access pricing regimes

A number of access pricing regimes have been used, or proposed for use, in Australia and overseas. The following are the main contenders.

- *Short-run marginal cost (SRMC)*: In the short run capacity cannot be varied. The access price is based on the change in variable (operating) cost from the greater use of the facility flowing from access provision. There is no element in the access price for the coverage of capacity costs, although SRMC could be defined to include the cost of any capacity specifically installed to facilitate access (Kahn and Shew 1987).
- *Long-run marginal cost (LRMC)*: In the long run all inputs are variable, so long-run marginal cost is the change in both operating and capital costs arising from a (permanent) increase in output. Different interpretations are possible.
- *Directly attributable incremental cost (DAIC)*: Under DAIC, capital and operating costs 'directly attributable' to the provision of access are included in the access price, inviting a more 'long-run' interpretation of marginal cost. As it actually operated, it included costs that were in no sense 'incremental'.
- *Total service (or element) long-run incremental cost (TSLRIC)*: The access price is based on the total allocable operating and capital costs of a service or production element, effectively averaged over the total number of units of use. It can also be thought of as the long-run avoidable cost of the element (per unit of output) reflecting the saving in operating and capital costs that would be made by the abandonment of the service.
- *Forward looking total service long-run incremental cost (FLTSLRIC)*: Like TSLRIC, but where the cost of providing access is based on using either the best-in-use or the best-available production technology, rather than the actual production technology in use.
- *Fully distributed (allocated) cost (FDC)*: Under this approach, the access seeker is charged the long-run marginal cost on a total service basis plus a share of 'common costs'. Common costs are allocated to different service areas on some basis such as proportionality to attributable costs. (Brown and Sibley (1986, pp. 44–51) is a good source on the traditional appeal of the FDC approach to monopoly telecommunications companies and their regulators.)
- *Efficient component pricing rule (ECPR) or the Baumol-Willig Rule*: This most famous of access pricing rules is where the access charge comprises long-run marginal cost on a total service basis (that is TSLRIC) plus the other aspects of the opportunity cost to the supplier of access to the essential facility, particularly the lost contribution to overhead costs and above-normal profits forgone (Baumol and Sidak 1994).

In 1996, Optus and Telstra renegotiated the basic access agreement, settling on 3.5 cents per minute, well in excess of the estimate of average cost of the LEN (2.5 cents per minute). There is also a separate interconnect agreement between Telstra and Optus Vision with a flat rate interconnection payment of around \$0.10 per local call originated on one network and completed on the other network. This was roughly equivalent to Telstra's average cost per *use* of the LEN.

5.4 Post-1997 access arrangements

Part XIC of the Trade Practices Act establishes an access regime specifically for telecommunications (ACCC 1997b). While related to Part IIIA, it incorporates characteristics argued to be specific to telecommunications. Once a network service is 'declared', it must be made generally available to access seekers unless specific exceptions apply. The ACCC may accept or reject access undertakings by carriers and, where commercial negotiations fail, arbitrate in access disputes.

The objectives of Part XIC revolve around the long-term interests of end users with the following features.

- The primary objective is the promotion of the long-term interests of end users of carriage services or of services provided by carriage services.
- Competition is an end in itself — the ACCC must consider the impact on competition in markets for telecommunication services.
- The objective of achieving any-to-any connectivity is seen as being in the long-term interests of end users.
- Encouraging the economically efficient use of, and investment in, telecommunications infrastructure includes considerations of technical feasibility, legitimate commercial interests of suppliers (including exploitation of economies of scale and scope) and investment incentives.

Declaration of carriage services relates to services that were available before 1 July 1997, declarations recommended by the Telecommunications Access Forum (TAF) or after an ACCC public inquiry. The TAF represents the interests of both access providers and access seekers.

With respect to access pricing (see ACCC 1997a), the ACCC focuses on an interpretation of the minimum costs an efficient firm will incur in the long run in providing the service as the basis for considering access prices. Effectively this means acceptance of total service long-run incremental cost (TSLRIC) on a 'forward-looking' basis (box 5.1). Further, the Minister may intervene with

guidelines to assist the access arbitration process. These are binding and no access undertaking or agreement can be incompatible with them.

There have been a number of inquiries since 1997 including into roaming (ACCC 1998) and local call services (ACCC 1999), and a draft decision (ACCC 2000) of a ‘general’ access charge of 1.5 cents per minute (0.84 cents conveyance cost plus 0.69 cents access deficit contribution) has been issued this year. A determination on fixed-to-mobile charges will follow declaration last year.

5.5 Assessment

Meets the need to combat monopoly abuse

A regulatory regime was necessary to facilitate competition in telecommunications. As chronicled earlier, Telecom/Telstra was an incumbent with a dominant position and it had consistently priced (or sought to price) access well above costs and charged others more than it appeared to charge itself. Given this history, it was clear that commercial negotiation between the operator of the facility and access seekers of the terms and conditions (including pricing) of access was not going to lead to an efficient outcome. The parties needed recourse to independent assessment or arbitration of the prices charged based on legislated criteria, and this has been provided by the legislation.

A possible criticism of the approach is that it values competition *per se* and places disproportionate emphasis on benefits compared with costs. Whereas Part IIIA appears to take a more neutral approach, Part XIC takes an explicitly pro-competitive line where competition is an end in itself and where the benefits of allowing access are given prominence.

The idea that competition should be seen as a means of improving productive and allocative efficiency, rather than as a goal in itself, was made clear by the Hilmer Committee of Inquiry into National Competition Policy:

Competition policy is not about the pursuit of competition *per se*. Rather, it seeks to facilitate effective competition to promote efficiency and economic growth while accommodating situations where competition does not achieve efficiency or conflicts with other social objectives. (Hilmer 1993, p. xvi)

While this is not exactly the philosophy of Part XIC, it is difficult not to have some sympathy with the legislators’ attempt to favour competition in the light of the historical exercise of monopoly power in providing access.

Daly and Stoeckl (2000) argue that an analysis taking into account costs, as well as benefits, (ie, a cost-benefit analysis or CBA) is more appropriate, and is consistent with the traditional economic efficiency approach applied to microeconomic policy issues by bodies such as the Productivity Commission. However, while Part XIC places primary emphasis on the ‘long-term interests of end users’ and cost factors are not explicitly listed among the criteria, more general efficiency considerations do enter through the ‘efficient infrastructure’ objectives.

Sound basic access pricing principle (TSLRIC)

Determination of an appropriate access price involves an important ‘balancing act’ between the price being ‘too high’ and ‘too low’. On the one hand, if the access price is set too high, access will be overly restricted with two types of efficiency cost. First, the owner of the facility would be allowed to exercise some degree of monopoly power (varying positively with the access price) in downstream markets, resulting in the usual efficiency costs of monopoly. Second, it would artificially encourage the creation of alternative facilities by potential competitors, resulting in wasteful duplication. Conversely, if the access price is set too low, there will be inefficient entry into the downstream markets and the operator of the facility will not have appropriate incentives to maintain or expand the capacity of the facility being accessed.

FDC, ECPR³ and DAIC, as it was applied in the early period of competition in Australian telecommunications, would generally lead to excessive access pricing, while basing the access price on SRMC would — except where capacity is heavily constrained — be likely to lead to the problems of under-pricing (box 5.1).

The ACCC’s (1997a) adoption of an interpretation of TSLRIC as the basis for determining access prices avoids these extremes, but partially basing TSLRIC on forward-looking costs perhaps errs in the direction of under-pricing. To the extent that TSLRIC relates to costs as they could be, rather than as they are, it could result in the access provider not achieving full cost recovery. On the other hand, it provides an incentive for the incumbent access provider to adopt lower cost technology.

Inefficient to have different prices for the ‘same thing’

Currently, use of the local exchange network (LEN) is based on at least three different pricing regimes.

³ Tye (1994) and Albon (1994 and 1995) provide critiques of different aspects of the inefficiency of the ECPR.

First, there are the arrangements between Telstra and interconnecting carriers for reticulating their long-distance and international calls. These charges are loosely based on an estimate of TSLRIC. (There are separate arrangements between the carriers for other instances of access, including for local calls and digital cellular access.)

Second, while the circumstances are unclear, it is unlikely that Telstra pays/charges itself for interconnection of its long-distance and international calls into the LEN on the same basis as other carriers are charged. Indeed, there are some indications that it does not have any explicit mechanism for pricing its own use. Under its access pricing principles, the ACCC (1997a) compares negotiated access prices with the price an incumbent charges itself for the same service. However, the ACCC notes that its 'rules of thumb' do not involve comparing these prices with costs. So long as the different divisions of Telstra remain under a single CEO and Board of Directors, there is an incentive and the ability to manipulate the internal transfer price. (This problem is further considered below.)

Third, not only are the conditions of use different between Telstra and the other carriers using the LEN, those applying to subscriber local call use are different again, with subscriber calls being untimed for charging purposes and having no locational or peak/off-peak distinctions.

Efficient pricing of, and investment in, the local exchange network is unlikely to be compatible with this diversity of pricing regimes for use of the same network.

- While TSLRIC is the right basis for pricing access to the LEN, in keeping with the theory of second best, it is also the right basis for pricing all uses of the LEN, whether as a production component or as a direct service. Getting one price 'right' while another or others remain 'wrong' does not necessarily improve efficiency.
- The current inconsistency is likely to mean both inefficient use of the network and inappropriate signals for development of the LEN. It is likely that the peak-load capacity of the local exchange network is inefficiently high because of the absence of timed local call charging. On the other hand, excessive final prices for domestic and international long-distance calls would mean less than optimal demand for these uses of the LEN. The balance of these forces probably results in a greater-than-efficient capacity requirement. Even if by accident the balance of these two effects on the capacity of the network resulted in the efficient capacity level, there are still efficiency costs from the inefficient pricing of the components of demand.

The 'access deficit' should not be included

Because of political and regulatory constraints, Telstra has not been able to cover its costs of providing customer access from its customer access charges, especially those for residential access. The resulting 'access deficit' has been used as a factor in determining interconnect access prices for rivals wishing to use the LEN to provide competitive services. This practice introduces an inefficient element into the access regime.

Inflating the cost of interconnect access raises long-distance, international and fixed-to-mobile charges, all areas where there are substantial efficiency costs at the margin. (This flows from the large gaps between prices and relevant marginal costs interacting with highly elastic demands.) On the other hand, covering the access deficit through higher customer access charges would result in little or no marginal efficiency loss (Albon, Hardin and Dee 1997, pp. 96–7).

Customer access charges remained constant for over ten years, as Telstra did not take the opportunity to vary them by the small amounts allowed under the sub-cap. However, the latest round of price capping placed customer access and local call prices in the same sub-basket, allowing Telstra greater scope for efficiency-enhancing adjustments to prices, especially in light of competitive pressures on local call prices. Customer access charges were raised substantially this year, but a large access deficit remains.

Structural separation of Telstra should have been part of the reform process

Where — as in Australian telecommunications — the essential facility is preserved within a vertically-integrated structure, the utility has an incentive and some ability to preserve the monopoly of downstream markets associated with its control of the facility. Some form of vertical separation of the natural monopoly elements of production processes is a means of attempting to insulate downstream markets from the monopolisation of the services supplied by an essential facility. Structural separation enables regulatory effort to be concentrated on the naturally monopolistic segments of the industry, and simplifies the overhead cost allocation problem.

Suppose that Telstra's local network were separated off as a profit centre, ring-fenced and subject to cost-based internal transfer pricing rules. This would result in at least some advantage with respect to reducing anticompetitive conduct. Potential rivals would deal directly with the local network division, and there would be less reason for other divisions to be involved in those negotiations. The local network division would not benefit commercially by favouring other Telstra businesses over external ones, although the overall organisation still could. This means that the

reduction in anticompetitive behaviour would be related to the extent to which decentralisation of management and adequate internal transfer pricing was fully implemented.

In light of this, structural separation of ownership may be desirable, but there are considerations for and against this approach.

- As the interests of the organisation as a whole could conflict with those of the separate businesses, separation of ownership of the divisions may be necessary to achieve an efficient access outcome. Hilmer (1993) favoured this approach.
- Separation could provide internal benefits to the organisation — Telstra is Australia's largest commercial organisation under a single board, and there may be diseconomies of scale because it is 'too big'. On the other hand, there may be economies from having the functions performed together ('economies of scope'). King and Maddock (1996, pp. 88–91 and chapter 8) express some reservations about vertical separation of utilities because of the loss of economies of scope. For example, there may be economies from operating local and long-distance services together. However, the source of the cost savings from joint operation of these services is not apparent and economies of scope have not been important in past decision-making about Australian telecommunications, especially the decision to allow the entry of new carriers.
- Regulatory and institutional structures would have to be adapted to be consistent with disaggregation. Contractual relationships have to replace internal management relationships when organisations are divided. In particular, institutional structures should ensure that vertical integration is not re-established via contractual means or cross-ownership.
- Splitting a monopolistic vertical production chain into separately-owned upstream and downstream production units could lead to 'double marginalisation', resulting in a higher price and lower output than under single ownership. However, this is unlikely in the circumstances of Australian telecommunications where regulators have substantial knowledge of the cost structure for providing both local and long-distance services, and where competition is emerging at both levels.

Convergence has not eroded Telstra's monopoly substantially

The exclusivity of Telstra's fixed-line local network is declining for a variety of reasons, particularly the introduction of competition and the impact of technological change in establishing alternative means of producing Telstra's products. While these forces will eventually reduce the strength of the required regulatory regime for access, at this stage, there remains a need for continuation of a strong access regime.

Within the telecommunications industry itself, the advent of cellular mobile technology has led to some minor erosion of the strength of Telstra's local monopoly. The impact is only minor because mobile telecommunications is not a stand-alone system, as it relies heavily on interconnection with the local network for completion of both fixed-to-mobile and mobile-to-mobile calls (Albon, Hardin and Dee 1997, chapter 1). Control of the local network remains crucial. Further, the penetration of mobile telecommunications remains limited.

Broader technological change in the communications industries ('convergence') is also having an impact. The telecommunications, media and information technology industries were once almost totally separate from one another, offering distinct services with firms that were largely industry-specific. Regulatory structures were (and remain) industry-specific — 'in separate silos' with the *Broadcasting Services Act 1992* (Cwlth) having a very different focus from the telecommunications legislation (Albon and Papandrea 1998). Convergence is changing the environment. For example, broadcasters are moving into the provision of two-way services and the internet can be used to make telephone calls.

Again, the impact of this can be over-stated. The local network is still a crucial production component for these alternative services. The substitute products they provide are far from perfect substitutes for traditional telecommunications products and penetration of customer access is very limited.

5.6 Principles of efficient access provision

1. Regulation should be independent of the access provider

Self regulation is inappropriate in the current circumstances of Australian telecommunications. Both principle and practice suggest that, if access decisions were left to the incumbent, rivals would be faced with a prohibitive access charge. An independent and 'uncaptured' regulator is required to avoid this.

2. Regulation should be independent of political interference

The utility reform process in Australia has revealed a tension between the pursuit of efficiency and the Government's desire to control pricing and investment for political purposes. With respect to access, the regulator should be given clear legislative criteria focusing on the attainment of efficient outcomes.

3. Base all prices on TSLRIC

Pragmatism and principle both suggest total service long-run incremental cost (TSLRIC) is unambiguously the right concept for efficient telecommunications pricing, but there are two main problems. First, the basis of costing capital is controversial, particularly the use of ‘forward-looking costs’. Second, there is still a problem of recovering unallocable costs (overheads) not included in TSLRIC.

4. Make appropriate allowance for differences in peak and off-peak costs

Consistent with efficient public utility pricing principles, prices should reflect the differences in the costs of meeting peak and off-peak loads, where peak-load costs include the capital cost of equipment dedicated to meeting the peak load (Kahn and Shew 1987). This principle is important in telecommunications, where peak and off-peak loads can differ greatly and the consequences for capital requirements of inappropriate pricing are substantial.

5. Price all uses of LEN (access and direct use) on the same basis

The local exchange network (LEN) has many different uses, but the time-specific costs of those uses are the same and should be charged on the same basis. Consistent with the theory of second best, getting some prices right while others are not does not necessarily enhance efficiency.

6. Separate issues of access from the ‘access deficit’

Providing access to rivals is supposed to lead to more efficient pricing, not to entrench existing inefficiencies in the price structure. The access deficit should be addressed in the first-best manner by raising customer access prices to reflect at least TSLRIC, and not as an excuse for introducing or perpetuating pricing inefficiencies elsewhere.

7. Separate the access providing part of Telstra from the operational parts

The maintenance of the vertically-integrated structure and inappropriate internal transfer pricing of Telstra means it has a potential conflict of interest with respect to the way it deals with access seekers that are also product market competitors.

References

- ACCC (Australian Competition and Consumer Commission) 1997a, *Access Pricing Principles — Telecommunications Guide*, ACCC, Melbourne, July.
- 1997b, *Telecommunications Industry — ACCC Role: An Outline*, ACCC, Melbourne, October.
- 1998, *Public Inquiry into Domestic Intercarrier Roaming Declaration under Part XIC of the Trade Practices Act 1974*, ACCC, Melbourne, March.
- 1999, *Declaration of Local Telecommunications Services*, ACCC, Melbourne, July.
- 2000, ACCC issues final access undertaking decision, Media release, 10 July.
- Albon, R.P. 1994, 'Interconnect pricing: an analysis of the efficient component pricing rule', *Telecommunications Policy*, vol. 18, no. 5, pp. 414–20.
- 1995, 'William Baumol's contributions to public utility economics', *Agenda*, vol. 2, no. 4, pp. 489–93.
- , Hardin, A. and Dee, P.S. 1997, *Telecommunications Economics and Policy Issues*, Industry Commission Staff Information Paper, AGPS, Canberra.
- and Papandrea, F.G. 1998, *Media Regulation in Australia and the Public Interest*, Current Issues, Institute of Public Affairs, Melbourne.
- Alston, R. 1988, *The Telecommunications Statement — A Competitive Liberal Alternative: A Critical Assessment of the Australian Government's New Directions in Telecommunications*, September.
- Baumol, W.J. 1995, 'Modified regulation of telecommunications and the public-interest standard', in Bishop, M., Kay, J. and Mayer, C. (eds), *The Regulatory Challenge*, Oxford University Press, New York, pp. 254–82.
- and Bradford, D.F. 1970, 'Optimal departures from marginal cost pricing', *American Economic Review*, vol. 60, no. 3, pp. 265–83.
- , Panzar, J. and Willig, R. 1982, *Contestable Markets and the Theory of Industry Structure*, Harcourt Brace Jovanovich, San Diego.
- and Sidak, J. 1994, *Towards Competition in Local Telephony*, MIT Press, Cambridge, Massachusetts.
- Bollard, A.E. and Pickford, M. 1995, 'New Zealand's 'light-handed' approach to utility regulation', *Agenda*, vol. 2, no. 4, pp. 411–22.
- Brown, S.J. and Sibley, D.S. 1986, *The Theory of Public Utility Pricing*, Cambridge University Press, Cambridge.

BTCE (Bureau of Transport and Communications Economics) 1993, *International Telecommunications: An Australian Perspective*, Report no. 82, AGPS, Canberra.

— 1997, *Interconnection Pricing Principles: The Choice of Pricing Principles for Standard Access Obligations in the Proposed Telecommunications Access Regime in Australia*, Working Paper no. 30, BTCE, Canberra.

Commission of Enquiry into the Australian Post Office (Vernon Report) 1974, *Report*, Parliamentary Paper no. 275, Canberra.

Crew, M., Fernando, C. and Kleindorfer, P. 1995, 'The theory of peak-load pricing: a survey', *Journal of Regulatory Economics*, vol. 8, no. 3, pp. 215–48.

Daly, A. and Stoeckl, N. 2000, 'Declaration of telecommunications services', *Agenda*, vol. 7, no. 2, pp. 113–24.

Dupuit, J. 1844, 'On the measurement of the utility of public works', translated by Barback, R.H. 1952, *International Economic Papers*, vol. 2, pp. 83–110.

Harberger, A. 1971, 'Three basic postulates for applied welfare economics', *Journal of Economic Literature*, vol. 9, no 3, pp. 785–97.

Hilmer, F., Raynor, M. and Taperell, G. (The Independent Committee of Inquiry) 1993, *National Competition Policy*, AGPS, Canberra.

IAC (Industries Assistance Commission) 1989, *International Trade in Services*, Report no. 418, AGPS, Canberra.

Kahn, A. and Shew, W. 1987, 'Current issues in telecommunications regulation: pricing', *Yale Journal of Regulation*, vol. 4, no. 2, pp. 191–256.

King, S. 1995, 'Guaranteeing access to essential infrastructure', *Agenda*, vol. 2, no. 4, pp. 423–31.

— 1997, 'National competition policy', *Economic Record*, vol. 73, no. 222, pp. 270–84.

— and Maddock, R. 1996, *Unlocking the Infrastructure: The Reform of Public Utilities in Australia*, Allen & Unwin, Sydney.

Lipsey, R.G. and Lancaster, K. 1956–57, 'The general theory of second best', *Review of Economic Studies*, vol. 24, pp. 11–32.

NCC (National Competition Council) 1997, *Competitive Neutrality Reform: Issues in Implementing Clause 3 of the Competition Principles Agreement*, AGPS, Canberra.

PC (Productivity Commission) 2000, *Broadcasting*, Report no. 11, AusInfo, Canberra.

-
- Pickford, M. 1996, 'Pricing access to essential facilities', *Agenda*, vol. 3, no. 2, pp. 165–76.
- PSA (Prices Surveillance Authority) 1989, *Inquiry in Relation to the Supply of International Telephone Services by the Overseas Telecommunications Commission*, Report no. 20, PSA, Melbourne.
- Quiggin, J. 1996, *Great Expectations: Microeconomic Reform in Australia*, Allen & Unwin, Sydney.
- Sherman, R. 1989, *The Regulation of Monopoly*, Cambridge University Press, Cambridge.
- Tye, W. 1994, 'The pricing of inputs sold to competitors: a response', *Yale Journal of Regulation*, vol. 12, no. 1, pp. 203–24.
- Vickers, J. and Yarrow, G. 1988, *Privatization: An Economic Analysis*, MIT Press, Cambridge, Massachusetts.
- Williamson, O.E. 1966, 'Peak-load pricing and optimal capacity under indivisibility constraints', *American Economic Review*, vol. 56, no. 4, pp. 810–27.

Discussant — *Rod Shogren*

Australian Competition and Consumer Commission

There is very little to disagree with in Robert Albon's paper. In those circumstances, the options for a commentator are to find a few small points of disagreement, to talk about the issues in the paper that they strongly agree with, or to talk about something completely different. This discussion addresses all three options.

D5.1 A comparison with the airlines industry

Consider the airline industry. Not as it looks now, but with some hypothetical differences to reflect the telecommunications industry.

Suppose Qantas owned the major airports in Australia. Qantas would allocate space at check-in counters in airports — the check-in counters that it used and the check-in counters used by competing airlines. Qantas would allocate time slots for take-offs and landings — time slots that it used and time slots used by competing airlines. Qantas would set the charges for aircraft movements — charges for its aircraft and charges for the aircraft of competing airlines. Would such arrangements engender confidence in the level of competition in the domestic airline industry in this situation?

Suppose it was said: 'The competing airlines should stop complaining. If they are worried about their access to airports, they should go and build some airports of their own'. It is not that easy to build airports or to jump the necessary regulatory hurdles, there are costs associated with complying with environmental impact statements and local government requests. Furthermore, there may not be a need for new airports. The current airport may not necessarily be working at full capacity, simply because competing airlines are having difficulty gaining access.

In this case, the problem for competition is ingrained market power. The power of the enterprise that owns the airports and the airline comes from the fact that airlines need access to airports. If Qantas owns the airport and a major airline, then Qantas has a strong incentive to use its power in airports to help its own airline, to the disadvantage of competing airlines. It could cross subsidise its airline operations, finding this to be a profitable strategy for its overall interests, and it could create obstacles for other airlines.

The analogy here is with the telecommunications industry. Competitors in telecommunications need access to Telstra's network to compete against Telstra in retail markets. Telstra is vertically integrated and comes from a statutory monopoly

position. It holds a very dominant position in most markets, and, in particular, the market for access to customers.

There is a need for a strong access regime that deals with the specifics of the telecommunications industry. The degree to which this follows, depends on how good the analogy is between the telecommunications industry and the hypothetical airlines industry just described.

One major difference between the two industries is the rate of technological change. There is the potential to allow new ways to compete for access to customers in the telecommunications industry. This means a telecommunications access regime needs to be able to withdraw from regulation, as that becomes appropriate. Note that the only two public inquiry processes that the ACCC currently has under way, regarding the declaration of telecommunications services, are both about narrowing the boundaries of regulation.

Returning to the airport analogy, could any regulator handle the situation such that competing airlines were on a genuinely competitive footing with the airline that owned the airports? Obviously, the first best solution for such an industry structure would be to deal with the structure directly. Unfortunately, this option for telecommunications has been ruled out in Australia.

D5.2 The telecommunications access regime

In his paper, Robert Albon suggests that the telecommunications access regime perhaps values competition and places disproportionate emphasis on benefits compared with costs. The test applied by the ACCC in creating rights of access and obligations to provide access is whether these arrangements are in the long-term interest of users. If this was interpreted exclusively to mean the promotion of competition, then Robert Albon's point would be correct. However, this is not what the test specifies.

The ACCC is also required to consider the efficient use of, and investment in, telecommunications infrastructure. Overpromoting competition can only mean promoting inefficient entry to the detriment of efficient investment and efficient use of infrastructure. The test itself, and the way it is interpreted, provides a balance between the benefits of competition and the costs that could be associated with wasteful competition.

In developing the principles and deciding that generally TSLRIC will be used, the ACCC's objective was to create the right balance between promoting competition and efficiency. In fact, the guide to declaration of telecommunication services

(ACCC 1999) is a straightforward attempt to devise a regulatory regime that would promote various types of economic efficiency.

It is interesting that in other countries there has been a greater emphasis on promoting competition, even at the expense of inefficient investment. This used to be the case in the United Kingdom. Policy was directed towards entry of new competitors, although this could lead to the inefficient duplication of facilities. This is not how the Australian regulatory regime has been established.

D5.3 Access deficit

Finally, the access deficit has been created by the fact that Telstra currently recovers, in its line-related revenues, considerably less than its line-related costs. Therefore, it has to make up the difference through the costs of calls and access charges to competitors.

The problem of access deficits is a considerable problem for the ACCC when setting access prices, the prices for local call resale and possibly even in setting prices for the unbundled local loop. The problem is due to price controls, which involve very large distortions in the market. The objectives served by price controls, that is, lower prices and guaranteed access for low income earners and low users of telecommunications services, could certainly be achieved in better ways than at present.

Reference

ACCC (Australian Competition and Consumer Commission) 1999, *Telecommunications services — Declaration provisions: A Guide to the Declaration Provisions of Part XIC of the Trade Practices Act*, ACCC, Canberra, July.

General discussion

The discussion focused on:

- the regulation of telecommunications;
- the appropriateness of TSLRIC as a pricing regime; and
- the merits of structural separation.

The regulation of telecommunications

It was suggested that Robert Albon's paper did not address the important issues of what and how to regulate telecommunications. The participant referred to Stephen King's paper that addressed these questions (see chapter 4 in this volume).

On the question of what to regulate, there was concern that the scope of regulation currently applied to the telecommunications industry was not justified. Of course, there were parts of the telecommunications network that fit the airport analogy. However, other parts of the network, which were growing most rapidly, and which were most significant for the future, were also regulated. For example, it was noted that the current access regime regulated:

- the provision of local calls on a resale basis in central business districts, where the network was extensively duplicated and triplicated;
- pay TV, where there were many alternatives, including free-to-air TV; and
- intercapital transmission, where again the network had been duplicated and triplicated virtually at all points in Australia.

Given the high level of regulation in the telecommunications industry, the question of whether the market would do it better was posed. To illustrate, a participant referred to the mobile phone networks of Australia. There were seven mobile phone networks in Australia and the dominant player in the market was not Telstra.

The appropriateness of TSLRIC as a pricing regime

There was agreement from several participants that TSLRIC was not an effective access pricing principle. It was pointed out that TSLRIC was merely a form of forward-looking average cost. It was not the economic answer to any economic question. There was concern that TSLRIC was not robust, and made it difficult to predict outcomes. For example, attempts to have peak and off-peak prices distinguished within an interconnection pricing scheme were abandoned because of objections from the ACCC.

Robert Albon replied that Stephen King's suggestion to use multi-part pricing for access was impractical. If SRMC were used, prices would fluctuate as demand grew and capacity was reached. Prices would then decrease when new capacity was installed.

The merits of structural separation

It was also suggested that structural separation was not relevant today. Every country that had moved down this path, such as the United States, had since moved away from it. Technological progress meant that the costs of structural separation now greatly outweighed the benefits. In addition, once there was imperfect downstream competition fed by a structurally separated upstream firm, with access prices higher than the marginal cost of access, structural separation would always lead to higher downstream prices. This was because an integrated provider would have an incentive to sell more access since the regulated price was above cost, but structural separation would remove this aggressive player. So structural separation would increase downstream prices, unless it were accompanied by a lower access price.

It was recognised that structural separation was not on the political agenda and that it would be difficult to now separate Telstra. Nonetheless, Robert Albon believed it was a desirable objective to at least have some structural separation within Telstra.

6 Improving the regulation of international air services

Geraldine Gentle[†]
Productivity Commission

6.1 Introduction

For over half a century, international civil aviation has been governed by a system of bilateral air service arrangements (ASAs) between countries. There are now some 3000 ASAs worldwide; Australia has over 50. ASAs set out the terms and conditions under which airlines can fly internationally. Typically, they specify capacity, frequency, routes, cities, ownership provisions, safety certification, price approval processes and many other details.

Restrictions on trade in goods are generally applied uniformly to all trading partners that are members of the World Trade Organization (WTO). In contrast, trade in international air services is largely excluded from the WTO's General Agreement on Trade in Services (GATS) and is not subject to the most favoured nation (MFN) principle. Instead, it is regulated within the framework of bilateral agreements, under the aegis of a UN agency, the International Civil Aviation Organization (ICAO).

In contrast to trade in goods, which is generally allowed unless specifically restricted, trade in international air services is prohibited unless specifically allowed. Various 'freedoms of the air' allow airlines to fly over, to, from, beyond, within and between bilateral partners and other countries. Other 'positive' provisions of ASAs allow airlines to undertake specific activities, such as setting up an office, in the foreign bilateral partner country.

Few other industries are subject to such intense *economic* regulation, in addition to the understandably high level of technical and safety regulation. This economic

[†] This paper draws heavily on the Productivity Commission report *International Air Services* (PC 1998). Additional commentary is the responsibility of the author.

regulation is known as the ‘bilateral system’. Under it, pairs of governments determine, on a bilateral basis, which airlines can fly between and beyond them, on what routes, how often, how many seats they can offer and, sometimes, even at what price. They also influence corporate structures and finances, because under this bilateral system, every country is expected to have its own ‘flag’ carrier, domiciled and generally at least majority owned and controlled domestically. In the past, most international airlines were actually owned by their respective governments.

This regulatory system restricts the efficient operation of airlines around the globe, increasing costs and prices and inhibiting the development of new and better services to consumers and users, with adverse effects on user industries, such as tourism, and economies generally.

From today’s perspective, this bilateral approach seems a peculiar way to organise any industry, let alone a transport system that is inherently based on networks. If ever there was a service industry where better regulation was needed, international aviation would have to be a prime candidate.

Aviation has been growing rapidly in spite of the regulatory framework and the need for change has become apparent around the world. Privatisation of airlines, deregulation in domestic aviation, the rise of alliances and the unwieldy nature of such high intensity regulation are all putting pressure on the bilateral system that is cracking under the strain. However, so far it has been immune to multilateral liberalisation, although some governments and the industry itself have been pushing the regulatory envelope.

World trade in most other goods and services is being liberalised, both unilaterally and in a multilateral framework, producing demonstrable gains in economic welfare. A different system for international air services is increasingly difficult to justify.

This paper examines the effects of the bilateral system and possibilities for reform.

6.2 The nature of international production and trade in air services

Each city origin and destination pair for air services can be considered a separate ‘product’, with limited substitution possibilities with other city pairs. These products are geographically differentiated and heterogeneous. Further, the characteristics of production in the airline industry mean that few routes are likely to be served by a large number of airlines, even in the absence of formal restrictions

contained in the bilateral system. Small group oligopolies, duopolies or even monopolies are likely to predominate (particularly on Australia's long thin routes) and the costs of entry can be significant.

One country alone cannot produce international air services. An international flight requires inputs from both the origin and destination countries in the form of infrastructure and rights to exercise various freedoms of the air (such as rights to embark and disembark passengers and freight, and to take off and land). This means that at least two countries must allow an international air service to be produced. Governments exercising control over their own airspace are able to specify conditions under which production may occur. This is in contrast with goods and most other services, whereby a country is generally free to produce whatever it likes on a unilateral basis.

There are other services that also cannot be produced without cooperation between at least two countries, such as international telecommunications and postal services. Multilateral free trade agreements have been concluded for both of these services, with significant benefits. Thus, the nature of the product does not imply that international agreements need be bilateral, nor that they should be of a highly constraining nature.

For most countries, the production of particular goods and services is small relative to worldwide output. Producers assume they will be able to export freely if they can produce at or below the world price. However, whether it is possible to export what they produce will depend on the trade policies of other countries.

Freedom to produce and trade does not apply in international aviation. Given the bilateral nature of production of any particular international flight, each origin and destination country has power to veto production. Without agreement, one country's airline cannot produce international services even for the country's own citizens, regardless of its relative efficiency.

The bilateral relationships that have evolved for international air services mean that each bilateral partner has considerable bargaining power. Each country can determine the size of the total supply to the bilateral market, not just its own level of output. Such market power is equivalent to the 'terms of trade' effect in international goods trade. Australia rarely possesses enough influence in world trade to enjoy bargaining power of such magnitude (wool and abalone are notable exceptions).

The bargaining power inherent in the bilateral system has allowed individual countries to pursue aviation policies that protect their carriers from international competition. The result is that airlines of varying levels of efficiency are able to

compete on the same routes, sustained by anticompetitive constraints in the ASAs and/or subsidies.

These assistance arrangements can be contrasted with assistance measures used in international trade in goods. For example, under the General Agreement on Trade in Services (GATS), the non-discriminatory MFN principle means that tariffs applied by WTO members are relatively transparent and generally apply uniformly to imports from all other WTO members. In contrast, the assistance provided through the bilateral system is extremely complex and opaque. It varies between every pair of countries, and in some cases the details are confidential.

The assistance has significantly muted competitive pressures and reduced the capacity of the most efficient airlines to drive out less efficient operators.

Without barriers to trade in international air services, airlines would need to compete on the basis of their efficiency relative to other carriers competing in the same markets — just as producers compete in other product markets. Consumers and other users of air services would benefit if more efficient carriers entered the market and enhanced competition. Australian airlines would also benefit from free trade in international air services if they were efficient enough to expand into markets for which access has been restricted.

6.3 Effects of the bilateral system

The economic regulations in the bilateral framework of international air services have had profound and largely deleterious effects on airlines and their users, which have been well documented by Findlay (1997), Tretheway and Oum (1992), Hanlon (1996) and others.¹ While liberalisation over the last decade has significantly lessened these effects, ASAs still affect where airlines fly, the number and frequency of flights they operate, what aircraft they use, their ownership and access to equity, and how much they can charge. They influence the nature of competition among airlines through controls on market entry, and the quantity and variety of rights allocated to particular airlines.

Airline costs and revenues

The regulatory system influences airlines' costs, sometimes facilitating the operation of efficient air services, but mostly increasing the costs of operation by restricting airlines' ability to achieve economies of scale and scope and to operate

¹ See chapter 6 of PC (1998) for a summary of these effects.

efficient networks. The constraints have implications for employment, and there are serious effects on the cost and availability of capital. Higher costs and restricted competition mean that prices are higher than they otherwise might be. The regulatory system also has implications for airlines' revenue and profitability.

The regulatory structure may affect the cost of labour. Entry limitations and capacity constraints have the potential to allow airlines to earn above normal returns, which may be appropriated by shareholders or paid out in higher than normal costs (including wages, salaries and working conditions).

Ownership and control restrictions prevent the merger and acquisition of international airlines across national boundaries. Consequently, airlines are constrained in achieving the efficiency benefits of operating a larger and potentially better integrated business. Growth through merger or acquisition enables airlines to achieve economies of scale and scope by consolidating airline functions. The merger of two airlines, for example, may allow them to consolidate their ground handling, maintenance, information technology and various managerial functions.

The restrictions may raise the cost and availability of equity capital, and distort choices between equity and debt capital, given that access to debt is not constrained by aviation regulation. Concerns about the cost-raising effects of the ownership and control requirements for national designation lie behind the proposal by the ICAO Secretariat to liberalise designation rules (ICAO 1994).

Restrictions on foreign investment in airlines and their ability to merge have prompted airlines to seek other ways of achieving the same benefits. The use of alliances and codesharing has grown in recent years as a means of minimising airline costs, extending market reach and increasing airline profitability. However, an alliance cannot achieve the same level of operational integration or cost savings possible within a single firm (IC 1997).

Many features of ASAs restrict the development of efficient airline networks. Capacity, frequency and route restrictions prevent airlines from developing their preferred range and quantity of services and from optimising their networks. To the extent that airlines cannot use the least cost combinations of aircraft types to carry passengers and freight, the costs of operating existing networks are higher than they otherwise might be. Further, they may be prevented from flying the optimum sized and configured network. Thus, there is the potential for costs to be reduced if airlines are able to operate the right aircraft at the right frequencies on an existing route. Airlines, by changing the design of a network and increasing its size, may also be able to decrease costs through economies of scale and scope.

The bilateral system assumed that point-to-point services would be the basis of international air services. While this may have been a reasonable assumption given the state of aircraft technology at the time of the Chicago Convention, it is not so today. This framework has hindered the development of efficient networks (including hub and spoke operations). Restrictions on rights to fly beyond a bilateral partner's ports, and the general requirement that flights must originate or finish in the home country prevent airlines from developing efficient networks based on overseas hubs.

The bilateral system of regulation involves significant compliance costs, obliging international airlines to devote considerable resources to tracking negotiations, liaising with government agencies and participating as technical advisers on delegations. ASAs that require periodic renegotiation impose higher compliance costs on the industry and are more costly to regulate than more liberal arrangements where prior agreement allows for further increases in capacity and other rights, or where there are no restrictions at all. Ownership and control restrictions also impose costs on airlines in monitoring their compliance with foreign ownership constraints, including the cost of monitoring share registers.

Some regulatory restrictions act to decrease the potential revenue from operating certain services. Even if a carrier is technically efficient, regulatory restrictions may restrict its ability to pick up freight and passengers at different points in the network. Restrictions on landing rights in particular cities, beyond rights, cabotage and stopover rights may all constrain the ability of airlines to optimise passenger and freight loads over different stages in their networks.

If airlines are able to land in certain places to discharge passengers and cargo, but are unable to take on further passengers or cargo for a subsequent stage of a round trip, they will be discouraged from operating the second stage.

Competition among airlines

The regulatory structure inhibits competition in many ways. It can prevent or deter entry, constrain capacity and limit the potential for airlines to win market share. Most ASAs also contain price regulation in one form or another. The degree of competition varies widely across routes. To the extent that competition would occur in the absence of regulation, restrictions decrease economic efficiency by raising the price of air travel, reducing the range of services offered, and discouraging innovation and cost minimisation.

Economies of scope and traffic density favour large airlines operating many services. On the demand side, a single carrier operating a long thin route with

multiple frequencies will attract better business than multiple carriers who each operate one service per week. Thus, markets tend to be concentrated with a small numbers of carriers operating on most routes. This applies even within the US domestic market, where many routes are highly concentrated despite the existence of a ‘competitive fringe’ of airlines capable of entry (Forsyth 1998).

The bilateral system of regulation contains numerous barriers to entry. Capacity, frequency, route, networking and investment constraints, and restrictions on codesharing can all be important. Furthermore, the continued use of substantial government subsidies is an obstacle to achieving economically efficient air services and has important implications for competition.

Capacity is used as negotiating coin and is still restricted in most bilateral agreements. Capacity constraints are credible only if they are binding or potentially binding. Capacity constraints inhibit the growth of the market, the airlines of the countries concerned and downstream industries, such as tourism. If third country carriers have capacity on such constrained routes, consumers and tourists may not be greatly disadvantaged, although prices are likely to be higher and choice constrained. The airlines of the bilateral partners may lose market share if they are restrained by capacity limits from providing extra services to meet growing demand.

Codesharing has the potential to promote competition among carriers on ‘thin’ routes where it may not be commercially sustainable for multiple carriers to operate services. Codesharing may provide a means for carriers to start-up services and test new routes. Codesharing may be an attractive option to the operating carrier in these circumstances when the additional traffic generated by the non-operating carrier lowers operating costs and makes it feasible to operate more services.

The inclusion of codesharing rights in ASAs allows airlines flexibility in terms of how they will contest markets. Thus, they have the potential to generate greater competition and to minimise the costs of operating on certain routes and the overall network. The omission of these rights from ASAs restricts the scope for foreign carriers to enter new markets and to minimise costs.

The impact of codesharing on competition will depend on the extent to which the carriers operating on a codeshare basis compete with each other on the same route or use codesharing to connect their services. Where carriers are already operating on the same route, and have relatively high market shares, codesharing is likely to reduce competition.

The development and use of computer reservation systems became an important competition issue in the late 1980s. An airline, by developing its own system which

travel agents and other airlines could then use, gained a competitive advantage by ensuring that its flights were listed more prominently on computer displays used in booking tickets. This issue became such a concern that it is now specifically addressed in ASAs, and is covered under the GATS.

Market concentration

The bilateral nature of regulation means that the designated airlines of the partner countries often dominate the air services between the two countries. In many cases, this means that a small number of airlines supply most of the direct services. Bilateral arrangements, by generally constraining the potential for entry and the ability of designated carriers of one country to take market share from carriers of the other country, help preserve high numbers of airlines at the global level, but relatively high market concentration on a country pair basis.

Airfares and services

Competition between airlines has been shown to result in lower prices, and an increased range and quality of services. Given the considerable differences among groups of customers, and the abilities of airlines to differentiate their prices, the effects of competition are likely to be different from one group to the next. Competition for price-sensitive travellers, such as people travelling for leisure reasons, is likely to lead to lower prices, whereas competition for price-insensitive travellers, such as business people, could focus on non-price aspects, such as frequency of service, in-flight services, and the use of airport lounges. Empirical research confirms this behaviour.

Regulations that lead to higher inbound international airfares reduce the number of inbound foreign tourists. Higher airfares also reduce tourists' expenditure on accommodation, shopping, restaurant meals, guided tours and internal transport. However, higher outbound international airfares encourage residents to travel domestically instead of overseas, benefiting local tourism operators. Thus, they encourage import substitution — that is, the supply of domestic rather than imported tourism.

Regulations that lead to higher airfares are a tax on exports of tourism services. Higher airfares sustainable under constrained ASAs impose net costs on tourism operators, as well as on Australian tourists. Higher airfares impede the development of inbound foreign tourism, and the tourism choices of Australians, more than offsetting any benefits to tourism operators from more Australians travelling at home instead of overseas.

Regulations that restrict the number of airlines serving particular routes, and the extent to which they can compete with each other, may limit the range of service levels — safety, ticket conditions, check-in and on-board facilities — from which passengers can choose. Capacity and frequency restrictions may constrain the introduction and expansion of direct flights, the timing and choice of services, and the ease and speed of connections. Passengers incur costs of less direct services in the form of increased travel time, lower availability of on-demand seats and less convenience, while tourism operators incur costs in the form of lower levels of inbound tourism. Lower levels of flight frequency reduce the opportunity of passengers to satisfy their itineraries and seat requirements, and reduce the potential for airlines to draw on the latent demand for air services.

More convenient flight scheduling has been found to be important to business travellers (Morrison and Winston 1989). Restrictions that affect the availability of services can be an important issue in the competition for discretionary travel by potential tourists.

The regulation of passenger services through ASAs can have implications for the supply and configuration of international air freight services. As most ASAs have been negotiated with passengers in mind, they do not always cater for the requirements of cargo operators. The nature of air freight makes for a very different market from passenger services. Freight users are not particularly sensitive about the route taken to deliver their freight and may accept less direct routes in exchange for lower rates. This gives them a wider range of choices over airlines and routes than passengers who are interested in relatively direct routes which conserve time. Competition is therefore quite vigorous, even within regulated markets. However, the constraints on network development affect the ability of specialist freight airlines to develop hubs outside their own country.

Economywide effects

Liberalisation of trade in goods and services has been a major factor in world economic growth over the past half century. There is nothing unique about production and trade in air services that would indicate that the substantial costs to airlines and their customers of constraints on activity and competition within the bilateral system do not have harmful effects on countries generally.

Liberalisation of trade in international air services is likely to bring substantial benefits to consumers, tourism and the other industries that rely on international aviation. It is also likely to benefit efficient airlines. Those countries with efficient airline industries would reallocate more resources to the airline industry and away from those industries that are relatively less efficient. Conversely, those countries

with relatively inefficient airlines would move resources out of airlines and into activities to which they were better suited, leading to economic gains in the long term.

6.4 What can be done to improve it?

The liberalisation agenda

The constraints on competition and trade in the bilateral regulation of international air services have significant, largely adverse, effects on airlines, users and the economy. The introduction of greater competition has led to substantial economic welfare gains, not just in Australia, but also overseas. On balance, liberalisation of trade and investment in international air services is likely to bring substantial benefits to consumers, tourism and other industries reliant on international aviation, as well as more efficient airlines.

In addition, world trade in most other goods and services is increasingly being liberalised in a multilateral framework. This is producing demonstrable gains to economic welfare and making it difficult to justify a different system for international air services. The APEC commitment to achieve free and open trade and investment for industrialised economies by 2010, and for developing economies by 2020, may also have important implications for regulating international air services. Removing restrictions on trade and investment conflicts with existing ownership and effective control restrictions that lie at the heart of the bilateral system of national designation.

Despite these pressures, the bilateral system has proven to be quite resilient — it incorporates technical regulation relating to international aviation procedures, safety and security, as well as economic regulation. While all of these types of regulation are mixed together within the bilateral system, it can accommodate more liberal economic regulatory arrangements without relaxation of technical and safety regulation.

However, bilateral liberalisation within the system takes time, is limited in effect and is piecemeal. Pressure will continue to mount from the global push for more liberal aviation agreements. Spurred by the United States, bilateral, reciprocal, ‘open skies’ agreements are spreading.

The global aviation industry has reached a critical point in its development. The inefficiencies imposed on the world aviation industry by the current bilateral system and the benefits of more efficient, internationally competitive air services are now

apparent. Reciprocal ‘open skies’ agreements are spreading. If the proposed trans-Atlantic common aviation area between the United States and the European Union eventuates, countries outside this group could be significantly disadvantaged.

Further liberalisation of international air services could include unilateral, bilateral, plurilateral and multilateral approaches.²

6.5 Ownership and control

Of all the bilateral system’s constraints on efficiency and competition, the most fundamental is the requirement that national flag carriers be locally owned and controlled. Because of such restrictions, nearly every country has its own international airline and an array of protective measures that support it.

Such restrictions on corporate structure and investment increase the costs to airlines. This reduces the ability for economies to specialise in the production of those goods and services in which they are relatively efficient and have a comparative advantage.³

Local ownership and control is not a necessary condition for the bilateral system to function. It is national designation, rather than ownership and control, that is necessary in order to assign rights to airlines, because they are not negotiated on a non-discriminatory basis and countries have different sets of rights.

Ownership and control provisions have been subject to much debate in recent years. At the 1994 World Wide Air Transport Conference, the ICAO Secretariat proposed that the rules for designation be liberalised. ICAO members indicated broad support for applying the ownership and effective control test at a regional level, for example, as Norway, Sweden and Denmark have designated Scandinavian Airline Systems. They also expressed in principle support for airlines being able to broaden their potential sources of investment by increasing limits on foreign investment in national carriers. However, they did not widely support the option of basing

² There would appear to be no particular problems in establishing a plurilateral or multilateral agreement that is consistent with the Chicago Convention. Several Articles in the Chicago Convention govern market access, the most important being Article 6, which requires ‘... special permission or other authorization of a contracting state in order for a scheduled international air service to operate into that State’. But ICAO has noted that ‘... the Article does not specify the process by which that permission or its terms may be given: permission could be exchanged bilaterally or multilaterally or even granted unilaterally’ (ICAO 1994, p. 2).

³ Restrictions on foreign ownership would also appear to be inconsistent with the APEC commitment to free and open trade and investment among developed country members by 2010 (and among developing members by 2020).

designation on ‘headquarters, central administration or principal place of business’ (ICAO 1994).

The lack of agreement among ICAO members resulted in no change to designation arrangements. In part, this highlights the difficulties of achieving change in multilateral fora such as ICAO. Many ICAO members were not convinced of the benefits of a more competitive framework for international aviation that reduced protection for their own national airlines. One of the major concerns was that relaxation of ownership rules might lead to unacceptable market access by third country carriers. This argument goes to the centre of the ownership debate — it is about protecting national airlines from competition and preventing the development of more efficient airlines that could, in a more liberalised market, be designated by a number of countries. Fear of the anticompetitive effects of the growth of ‘mega-carriers’, principally US carriers, was also apparent.

Airlines are now finding ways of overcoming some of these restrictions, for example by forming alliances or establishing contorted ownership arrangements.

The 1994 World Wide Air Transport Conference also highlighted concerns that liberalisation of the ownership criteria could compromise airline safety, particularly through flags of convenience (ICAO 1994).

These concerns largely reflect confusion between the concepts of ownership and designation. Ownership is not a necessary condition for designation, accountability or technical and safety regulation. However, designation is important for technical and safety regulatory purposes because countries are accountable for maintaining the safety standards of their designated carriers.

These non-economic regulatory functions of designation would remain, even if international air services were available on a non-discriminatory MFN basis. Some system of accountability between airlines and government regulators would continue to be required.

When there are clear lines of accountability between airlines and the countries designating them, it should not matter where they are owned, any more than in other industries. For example, national governments take responsibility for regulating the safety of pharmaceuticals through recognition of other countries’ certification processes as well as through their own — they do not need to own all pharmaceutical production and testing facilities in order to regulate safety. Indeed, in some situations, it might even be easier to enforce safety standards on airlines when the owners are foreigners. If a country is lax about enforcing universal, compulsory safety standards, ownership requirements for designation will not ensure that an airline is safe.

Although reform of ownership and control requirements for designation are critical for liberalisation of the bilateral system, options are limited because the provisions are actively monitored and enforced by various countries. Because designated airlines must be recognised by bilateral partners, it is risky for any individual country to change the criteria for designation of its airlines unilaterally. The United States showed that it was prepared to exercise its ‘muscle’ when it denied Aerolineas Argentinas status as an Argentinian carrier once it was no longer majority Argentinian-owned. This particular dispute was resolved at the price of concessions to the United States. It highlights the risks to countries from unilaterally changing designation criteria.

The current ownership and control requirements are inhibiting the development of a prosperous and sustainable international airline industry. Global capital markets are requiring higher performance standards from airlines and nations. There are simply too many poorly performing international airlines that are burdening national economies. Conversely, efficient airlines are denied the opportunity to develop to their full potential. If a country could designate an airline domiciled elsewhere, then it could improve its air services and free up scarce capital and technical resources for other, more productive uses.

Options for reform include basing designation on place of incorporation, principal place of business or other evidence of commitment to providing air services for the country. Ownership could be removed as a criterion for designation; regulation of foreign investment could be aligned with that for other industries.

New Zealand has relaxed the foreign ownership requirement in its more recent bilateral agreements, opting for the less restrictive ‘effective control’ requirement. In its submission to the Productivity Commission’s Inquiry into International Air Services, Air New Zealand noted that:

For the time being, airlines that complied only with these liberalised criteria would be unable to exercise traffic rights under bilateral agreements containing traditional ownership requirements. However, the intention of New Zealand and the bilateral partners with which it has negotiated these arrangements is to establish precedents that might increasingly be followed. (Air New Zealand 1998, p. 9)

As the recent financial crisis has focused attention on the need for capital to be used more efficiently, it may be opportune to try again to achieve multilateral agreement to liberalisation of the ownership and control requirements for national designation.

The Australian Government could join with other like-minded governments in a push for reconsideration by the members of ICAO of the 1994 proposals for liberalisation of the ownership and control criteria for national designation.

However, such multilateral action is unlikely to be swift. It might be useful also to pursue this initiative within APEC, whose members account for a substantial proportion of world aviation, or as part of plurilateral reform. In the meantime, countries should negotiate with bilateral partners to incorporate in their own ASAs a more liberal means of designating airlines which does not rely on ownership restrictions.

Another option to liberalise ownership and control criteria for national designation would be to develop regional approaches. Each country would still negotiate its own ASAs, but the airlines could be owned by citizens of any country in the group. While this would require changes to existing bilateral agreements, it should be possible to gain ICAO recognition of such a regional arrangement.

These are very much second-best options relative to liberal multilateral reform.

6.6 Unilateral liberalisation

Australia has undertaken most of its trade liberalisation on a unilateral basis. There are a number of good reasons for this:

- the trade barriers themselves were imposed unilaterally;
- the barriers may have benefited some producers, but the costs to consumers, users and the rest of the economy were quite high;
- bilateral deals on a reciprocal basis are precluded generally in areas of trade covered by the WTO, apart from customs unions and free trade areas; and
- Australia is a small country in international trade in most goods, so its bargaining power has been relatively weak in most multilateral trade negotiations.

These factors meant that Australia was unlikely to be able to negotiate better arrangements by waiting for multilateral negotiations to conclude, but would continue to bear the self-imposed costs for the duration of the negotiations. In these circumstances, it has been clearly in Australia's interests to liberalise unilaterally. Would it be equally advantageous for Australia or any other country to pursue unilateral liberalisation of international air services while the bilateral system remained?

One option would be to grant unilaterally to other countries unlimited rights to supply international air services into and out of Australia, without requiring any extra rights for Australia in return. This option would mean that Australia would cease to negotiate updated or new bilateral agreements and would rely on the rights

currently available for Australian carriers. Without expecting any reciprocal action by other countries, Australia would open its skies to foreign carriers, unilaterally removing all constraints on traffic and access rights.

Would such an arrangement improve efficiency in resource allocation in Australia and provide Australia with more efficient international air services? Would consumers and other users have a greater range and quality of services at lower prices?

Effects on airlines

In the absence of the bilateral structure, if all countries adopted unilateral ‘open skies’ approaches, the structure of the global airline industry would change radically. Efficient carriers would displace inefficient carriers. Users of air services would benefit from more competitive fares.

Unilateral ‘open skies’ within the bilateral structure would not produce the same outcomes. For the purposes of this discussion, it is assumed that no other country reciprocates Australia’s unilateral open skies policy. The only international rights that would not become frozen in time would be those where capacity is not restricted, such as in the ASAs with the United States and New Zealand. However, traffic beyond these countries is constrained, so it is likely that Australia’s beyond traffic would continue to be limited, while other countries’ carriers would not be limited in exercising beyond rights from Australia.

By allowing a bilateral partner to have unconstrained access, the airlines of that country stand to gain an increasing share of traffic to and from Australia. However, Australian carriers would not be able to expand their services elsewhere, as the rights for Australia’s airlines would be fixed for as long as the bilateral system remained entrenched.

Australian airlines would be severely disadvantaged by a policy of unilateral ‘open skies’. They would be unlikely to share in market growth with increased capacity and frequency (except in the United States and New Zealand). Further, they would not be able to expand their networks to include new destinations.

Consumer effects

Australian users of air services would benefit from a policy of unilateral ‘open skies’ if it led to a greater quantity and range of services and lower prices, or if foreign carriers were more efficient than Australian airlines and generated effective competition.

However, there is no guarantee that these changes would occur while the bilateral framework continued elsewhere. Without opportunities for growth, Australian airlines would lose market share over time. Australian airlines would be denied the market growth that they would otherwise have been able to contest. Consumers who prefer to use Australian airlines, but are no longer able to do so, would also bear the costs of having to use less efficient or less preferred carriers. As a result, both Australian producers and consumers would incur costs. It would also be difficult for Australia to negotiate any new beyond rights.

The quantity and range of flights from Australia to foreign destinations would only increase if the relevant governments permitted this to occur. With the bilateral framework still in place, Australia's bilateral partners would continue to control capacity and frequency on Australian routes. They would be able to control entry and curtail competition not only by their own carriers, but by third countries' carriers as well.

It is possible that, if Australian airlines were no longer an effective competitive force, airlines from bilateral partner countries would reduce the level of service and increase prices. Australia could retaliate by retreating from its unilateral liberalisation, but would be unlikely to regain the conditions it relinquished. The damage to the Australian airlines could be substantial, for little benefit to consumers. While Australia might be tempted to retaliate against an individual country in other areas of trade, the scope for such action is limited. Most of Australia's trade in goods and services is covered by non-discriminatory MFN principles under the WTO.

As long as air services are regulated in a bilateral framework, the degree to which Australia would benefit from a policy of unilateral 'open skies' would depend entirely on how its partners behaved in the new environment. Under some circumstances, Australia could be worse off than it is now, and opportunities for gain in the future could be forgone.

6.7 Bilateral liberalisation

Many countries, including Australia, have been pursuing a policy of gradual liberalisation on a bilateral basis. Substantial further liberalisation can be achieved.

Reciprocal 'open skies' agreements

One option for the liberalisation of international air services would be to negotiate bilateral 'open skies' agreements on a reciprocal basis.

The term ‘open skies’ does not have a precise meaning, but is commonly used to refer to the US template for the liberalisation of ASAs. These agreements generally remove restrictions on traffic to, from and beyond the bilateral partner, as well as regulation of airfares and the number of designated carriers. Cabotage rights are not negotiable and the limits on foreign ownership of US carriers remain highly restricted.

The aim of a reciprocal ‘open skies’ policy would be to remove restrictions in ASAs on competition and trade between the airlines of each of Australia’s bilateral partners, consistent with technical and safety regulation. Negotiations for reciprocal ‘open skies’ agreements should start with everything on the table. In addition, a reciprocal ‘open skies’ agreement would incorporate restrictions on government subsidies to airlines where necessary.

While the basic constraints of the bilateral system on airline networking and organisation would remain, open skies agreements would at least enable airlines to respond quickly to market changes and new opportunities. Consumers would benefit from greater capacity, expanded networks and increased competition as carriers take up opportunities to serve new markets. It would also minimise administrative and compliance costs for governments and airlines associated with negotiating incremental changes to ASAs.

The US ‘open skies’ policy is often criticised for not being a very ‘open’ and liberal policy by virtue of it retaining restrictions on cabotage and ownership and control. The New Zealand ‘open skies’ policy provides a more liberal approach to foreign ownership than the United States, as well as allowing for the negotiation of cabotage rights. Following the Australian Government’s acceptance of the Productivity Commission’s recommendation that Australia negotiate reciprocal ‘open skies’ agreements, some progress has been made, but the process is very slow.

Switching ASAs from ‘positive’ to ‘negative’ lists

One reason for the slow progress of reform lies in the framework of ASAs themselves. The ‘positive list’ of activities that an airline is allowed to undertake inhibits airlines from activities for which there is no specific provision. Airlines often require their governments to intercede for them before they can develop new or expanded services. In most other industries, international traders are able to do anything that is not constrained by a ‘negative list’ of specific restrictions. This gives companies much greater freedom to do business, and importantly, to grow and innovate.

Bilateral reform, including reciprocal ‘open skies’ agreements, would be enhanced if the ASAs were to shift to a ‘negative listing’ approach. This would allow carriers from each country to conduct air services without restriction, except for items that were specifically listed as not being allowed. Any restrictions under the agreement would be negotiated exceptions.

The negative listing approach has a number of advantages. First, it enables any new form of trade to be automatically included as part of the trading arrangements. For example, the widespread introduction of codesharing by airlines around the world meant that most ASAs had to be renegotiated to include codesharing specifically. Second, it is generally easier to negotiate the removal of remaining restrictions than to negotiate for the addition of certain activities or traffic rights. Third, it is more consistent with the General Agreement on Trade in Services (GATS) trading framework and would therefore facilitate a transition to a broader plurilateral or multilateral trading system for air services.

6.8 Plurilateral open club

As reciprocal ‘open skies’ agreements become increasingly common on a bilateral basis, the possibility arises for the development of a common agreement among a number of countries that would extend the scope from a bilateral to a plurilateral context. This could involve the creation of an open club of countries that would provide the benefits of the common liberal ‘open skies’ agreement to all who joined.

Characteristics of the club

The current regulatory framework has demonstrated its ability to accommodate a degree of liberalisation. However, it still inherently limits the potential for competition and further development of the international aviation industry.

A more promising option for reform may be to develop an aviation open club of countries based on a common agreement that would allow all carriers within the group to have similar sets of rights. For example, Australia might start a club with countries X and Y which gives the carriers of all three countries unlimited rights to operate into and out of any international airports of all three countries. Consideration might also be given to granting cabotage rights within the club. A key feature of the club is that it would be open to other countries to join on the same conditions.

Compared with a more general multilateral approach (where the potential for reform could be limited to what the least restrictive country would be prepared to negotiate), the open club offers a better chance for initiating worthwhile reform. This is because it could evolve from a small group of like-minded countries that are prepared to liberalise their aviation sectors. A common agreement would also have a distinct advantage over the complex set of bilateral agreements it would replace.

The benefits to members would grow in proportion to the size of the network. At the same time, the opportunity costs of not being a member of the club would also increase. Non-members run the risk of becoming isolated by air traffic that responds to more liberal conditions available within the club. Countries that may be reluctant to negotiate ‘open skies’ agreements on a bilateral basis may find they have little choice but to join the club.

The open club idea draws on the principles of open regionalism, but with an important difference. Under the WTO, signatories are required to treat the products and services of other countries on a non-discriminatory MFN basis — that is, in general, they cannot treat the products of one foreign country differently from those of another, except in a free trade area or a customs union. Open regionalism describes the action of a group of countries reducing impediments to trade among themselves, but not creating any new artificial distinctions or discrimination between members and non-members (Elek 1996). For example, APEC commitments to trade reform amount to a coordinated program of unilateral actions by members to reduce trade barriers without discriminating against non-members.

An aviation open club would operate differently. International air transport (except for certain ground-based services) has been specifically exempted from the GATS. Most countries apparently wish to retain the bilateral system, and that system fundamentally contravenes the non-discriminatory MFN principle. An aviation open club cannot help but discriminate against non-members because it would grow out of the bilateral system. The difference is that club members would offer one set of rights to all other club members and other sets of rights (bilaterally) to non-members. But non-members would be free to join. As membership grew, discrimination between nations would decline. Transparency would be important to ensure that non-members could understand the conditions of membership, and be forewarned about changes to club rules.

The difficulties of establishing such a club should not be underestimated. Not all countries may wish to join a club, or allow certain other countries to do so. Countries with entrenched political or other differences may not wish to cooperate in such a way. Airlines in member countries may have an incentive to resist extending the boundaries of the club if it would mean entry of other more competitive airlines. However, accession would be important because the overall

benefits, including benefits to users, would grow as the size of the aviation network grew.

The incentive for some countries to join would relate to the benefits and costs from better access to existing members. An open club would enhance the development of efficient networks by giving all members what are effectively matching sets of fifth and seventh freedoms. But some countries that already have important international hubs may perceive an ‘open skies’ policy as detrimental to their airlines unless it opens further opportunities for extending the airlines’ networks. In this more competitive framework, some carriers would be in a better position than others, and some rationalisation may be inevitable. The benefits of obtaining access to the network could help countries that may not have much to trade in a bilateral framework.

To enhance the chances of the open club becoming a vehicle for widespread reform, the club should:

- not intentionally disadvantage outsiders any more than the bilateral agreements it replaces;
- have transparent rules;
- allow new members to join on the same terms and conditions; and
- contain provisions relating to competition policy, particularly relating to restrictions on significant government subsidies.

A result comparable to an open club could be achieved by negotiating similar bilateral agreements between a series of partners. But each agreement would then need to contain similar provisions; this would be difficult to achieve through a series of bilateral negotiations, except for a large country like the United States that is able to pursue a uniform template in its ASAs.

The United States stands to benefit much more than its smaller bilateral partners from extending bilateral ‘open skies’ agreements. With a relatively small number of agreements, it is able to achieve substantial networking opportunities for its airlines that would not be open to the partners, unless they all concluded similar agreements with each other. For example, to offer the airlines of 30 bilateral ‘open skies’ partner countries the same possibilities for network development as enjoyed by the US carriers, there would have to be 465 bilateral ‘open skies’ agreements.

The same network possibilities would exist for all the airlines with a single plurilateral ‘open skies’ agreement covering the 31 countries. Progress would probably be best served by a small number of countries developing a plurilateral

open club agreement and inviting other countries to join, rather than trying to develop it through an extensive network of bilateral agreements.

Current proposals to turn the Atlantic into an EU-US lake could have severe effects on other countries if the arrangement is not open to outsiders. The challenge is to ensure that regional or plurilateral agreements are used as a stepping stone to global liberalisation and not as an exclusive device to favour members. Openness is a key factor in this.

6.9 Multilateral liberalisation

Regulating international air services in a liberal multilateral framework offers potential benefits over what might be achieved through other approaches. A liberal multilateral agreement which covers all or most countries would allow air services to develop in response to market pressures. Efficient carriers would replace inefficient carriers and the removal of regulatory barriers to entry would enhance competition. Geographic constraints, like Australia's position at the end of long haul routes, would no longer constrain airlines' ability to grow and develop services. A multilateral system would also be easier to administer and comply with than the current bilateral system. It would be consistent with the rest of the world trading system and could encourage liberalisation by more conservative nations.

In its submission to the Productivity Commission's Inquiry, the Australian Chamber of Commerce and Industry described what a multilateral agreement could include:

Key elements of any [multilateral] negotiations should include: ensuring liberal market access, with no restrictions (bar genuine safety considerations) on capacity or frequency of services; charges, prices and fares should be set by the market, subject only to bona fide competition policy issues; both passenger and freight movements; and, transparent and welcoming accession arrangements for countries which wish to join during or after the negotiation processes. (ACCI 1998, p. 5)

Attempts have been made to bring air services into a multilateral framework through ICAO and the WTO. A first, but unsuccessful, attempt was made at the Chicago Convention in 1944. Although ICAO has responsibility for developing economic regulation, most countries have remained wedded to the bilateral framework. ICAO has focused on developing standards for safety and security. Responding to the Uruguay Round of multilateral, multisectoral trade negotiations and the possibility that air services might be included in the GATS, the ICAO Secretariat proposed various reforms to economic regulation of international air transport. The ICAO assembly considered them in 1994, but did not agree to them.

The Uruguay Round also made little headway on air services. Air services are within the ambit of the GATS in principle, but an Annex explicitly excludes most of them. Only aircraft repair and maintenance, the marketing of air services and computer reservation systems are covered. This Annex can only be amended or removed by a two-thirds majority (currently 88) of WTO members. It provides for periodic reviews at least every five years, with the first review to take place by 2000.

If air services were to be covered by the general provisions of the GATS, ASAs between WTO members could become redundant. However, WTO members would need to maintain bilateral agreements with countries that are not members of the WTO.

Extending non-discriminatory, MFN treatment to air services would prohibit countries from discriminating among WTO members, except in the context of a trade agreement, such as a free trade area or a customs union. This absence of discrimination under the GATS means that the nationality of airlines and the national designation of airlines, as under the present bilateral system, would be less relevant. A country would still be able to restrict access to its airspace and to its domestic traffic, but it could not do so in a manner that discriminated among foreign airlines, except in the context of agreements for economic integration. Nor could the controllers of airspace, or the operators of airports, discriminate among foreign airlines on the basis of nationality.

However, given existing attitudes to air traffic rights, it would not be easy to obtain the two-thirds majority to repeal or amend the Annex, especially given that some otherwise liberally minded WTO members, such as the United States and the European Union, supported the exemption in the first place (DTRD 1998b). Indeed, as noted above, there are advantages to US carriers from preserving and extending the hub and spoke set of 'open skies' bilateral agreements around the US hub, rather than fostering a multilateral 'open skies' agreement.

As Havel (1997, pp. 405–6) noted:

A global multilateral instrument is currently unattainable. The bilateral system, centred in national sovereignty, is ideally suited to those states that continue to assert the well-being of their national carriers as the highest priority in aviation policy. Any attempt to corral these states into a wider multilateral framework would have GATT-like consequences: agreement may result, but it would be pitched to the 'lowest common denominator' ...

There are real dangers in substituting multilateral regulation for the current bilateral framework, if the agenda is not clearly focused on increasing efficiency, competition and consumer welfare.

References

- Air New Zealand 1998, Submission to the Productivity Commission Inquiry into International Air Services, Submission no. 6.
- Airline Business 2000a, 'Sans Frontiers', *Airline Business*, February, pp. 34–5.
- 2000b, 'Freedom's Path's', *Airline Business*, March, pp. 74–7.
- 2000c, 'Play by the Rules', *Airline Business*, April, pp. 72–4.
- Ansett Australia Limited and Ansett International Limited 1998, Submission to the Productivity Commission Inquiry into International Air Services, Submission no. 19.
- Australian Tourist Commission 1998, Submission to the Productivity Commission Inquiry into International Air Services, Submission no. 29.
- ACCI (Australian Chamber of Commerce and Industry) 1998, Submission to the Productivity Commission Inquiry into International Air Services, Submission no. 26.
- DTRD (Department of Transport and Regional Development) 1998a, Avstats Database, Canberra.
- 1998b, Submission to the Productivity Commission Inquiry into International Air Services, Submission no. 33.
- Dresner, M. and Tretheway, M. 1992, 'Modelling and testing the effect of market structure on price: the case of international air transport', *Journal of Transport Economics and Policy*, May, pp. 171–84.
- Elek, A. 1996, 'APEC: an open economics association in the Asia-Pacific region', in Bora, B. and Findlay, C. (eds), *Regional Integration and the Asia-Pacific*, Oxford University Press, Melbourne, pp. 223–34.
- Federal Aviation Administration 2000, *Beyond Bilateral Agreements*, Proceedings from the Aviation in the 21st Century — Beyond Open Skies Conference, 6–7 December, Washington DC.
- Findlay, C. 1997, 'The APEC Air Transport Schedule', *Pacific Economic Papers*, no. 273, Australia-Japan Research Centre, Australian National University, November.
- , Chia, L.S. and Singh, K. (eds) 1997, *Asia-Pacific Air Transport: Challenges and Policy Reforms*, Institute of South East Asian Studies, Singapore.
- and Kissling, C. 1997, 'Flying towards a Single Aviation Market across the Tasman', in Findlay, C., Chia, L.S. and Singh, K. (eds), *Asia-Pacific Air Transport: Challenges and Policy Reforms*, Institute of South East Asian Studies, Singapore, pp. 181–91.

-
- , Hufbauer, G.C. and Jaggi, G. 1996, 'Aviation reform in the Asia Pacific', in Hufbauer, G.C. and Findlay, C. (eds), *Flying High: Liberalizing Civil Aviation in the Asia Pacific*, Institute for International Economics, Australia-Japan Research Centre, Australian National University, Canberra.
- Forsyth, P. 1998, 'The gains from liberalisation of air transport: a review of reform', *Journal of Transport Economics and Policy*, vol. 32, no. 1, pp. 73–92.
- Gregan, T. and Johnson, M. 1999, *Impacts of Competition Enhancing Air Services Agreements: A Network Modelling Approach*, Productivity Commission Staff Research Paper, Ausinfo, Canberra.
- Hanlon, P. 1996, *Global Airlines: Competition in a Transnational Industry*, Butterworth Heinemann, Oxford.
- Havel, B.F. 1997, *In search of Open Skies: Law and Policy for a New Era in International Aviation*, Kluwer Law International, The Hague.
- IATA (International Air Transport Association) 2000, *Aviation Information and Research*, http://www.iata.org/air/air_htm (accessed 26 May 2000).
- IC (Industry Commission) 1997, *The Economic Impact of International Airline Alliances*, Information Paper, Australian Government Publishing Service, Canberra.
- ICAO (International Civil Aviation Organization) 1994, Liberalization of market access with a 'safety net' and full or progressive introduction, Report of the World Wide Air Transport Conference on International Air Transport Regulation: Present and Future, Montreal, 23 November–6 December.
- Jackson, G. 1998, Boost for the Regions — International access to regional airports to be liberalised, Press notice 461, Transport, Department of the Environment, Transport and the Regions, London, June.
- Morrison, S. and Winston, C. 1989, *Enhancing the Performance of the Deregulated Air Transportation System*, Brookings Institute, Washington DC.
- National Commission to Ensure a Strong and Competitive Airline Industry 1993, *Change, Challenge and Competition*, Report to the President and Congress.
- NCC (National Competition Council) 1997, *Applications for Declarations of Certain Airport Services at Sydney and Melbourne Airports*, Reasons for Decision, Melbourne, May.
- Ng, Y.K. 1987, 'Equity, efficiency and financial viability: public utility pricing with special reference to water supply', *Australian Economic Review*, 3rd quarter, pp. 21–35.

-
- OECD (Organisation for Economic Cooperation and Development) 1997, *The Future of International Air Transport Policy — Responding to Global Change*, Paris.
- Oum, T.H. and Yu, C. 1997, *Winning Airlines: Productivity and Cost Competitiveness of the World's Major Airlines*, Kluwer Academic Publishers, Massachusetts.
- PC (Productivity Commission) 1998, *International Air Services*, Report no. 2, Ausinfo, Canberra.
- Rose, N.L. 1990, 'Profitability and product quality: economic determinants of airline safety performance', *Journal of Political Economy*, vol. 98, no. 5, pp. 944–64.
- 1992, 'Fear of flying? Economic analyses of airline safety', *Journal of Economic Perspectives*, vol. 6, no. 2, Spring, pp. 75–94.
- Savage, S., Smith, D. and Street, J. 1994, The effect of competition on international air fares, Paper presented at the 23rd Conference of Economists, Gold Coast, Queensland, 25–28 September.
- Street, J., Smith, D. and Savage, S. 1994, 'An analysis of the trade-offs in international aviation rights', *Papers of the Australasian Transport Research Forum*, vol. 19, pp. 183–206.
- Tretheway, M. and Oum, T.H. 1992, *Airline Economics: Foundations for Strategy and Policy*, Centre for Transportation Studies, University of British Columbia, Vancouver.
- World Trade Organization 1998, *Air Transport Services: Background Note by the Secretariat*, Council for Trade in Services, Geneva.
- 1999, *Air Transport Services: Communication from New Zealand, Chile and Singapore*, Council for Trade in Services, Geneva.

Discussant — *Christopher Findlay*

Australian National University

This paper provides a thorough review of the nature of the regulatory system, its distinct characteristics and its impacts on airlines. It highlights the costs of the current arrangements to all parties, not just consumers.

D6.1 Assessing the current regulatory arrangements

International aviation services are subject to a complex regulatory system. To assess this regulatory system, it may be appropriate to apply the components of a GATS-type necessity test. There are several questions that can be asked.

Does this regulation apply to a legitimate policy objective?

The paper discusses the effects and directions for reform, but it is useful to consider the objectives of regulation, as it helps to define the reform program.

Gerry Gentle reviewed some possible objectives, but the system appears to have competitive objectives. Two countries must cooperate to provide international air services. The risk is that one will apply its market power to the detriment of others, based on its control of access to airspace. This was a genuine concern at the end of the war. Consequently, a series of bilateral agreements was constructed.

There is a general class of problem that applies more generally to network activities — where operation requires cooperation of both the sending and receiving party.

Is the current policy the most appropriate form of regulation?

If the objectives for regulation in aviation are justified, and that is debatable given the increasing density of the market, alternative regulatory structures must be considered. The obvious alternative is a policy targeted precisely at the objective, in this case access to air space or access to ground infrastructure.

These sorts of competitive issues and how they are handled are worth more emphasis than is given in Gerry Gentle's paper. Indeed, people in the industry talk about the capacity of the bilateral system to deal with problems, such as access to ground infrastructure, but are not keen for radical change. The bilateral system is awkward for this group, but provides them with a chance to capture rents. Alliances and codesharing remove some of the burdens of collecting those rents efficiently.

The current regulatory framework is clearly not the least trade restricting, as the review of its effects in Gerry Gentle's paper demonstrates. The irony is that a structure that can be interpreted as trying to solve a competition policy problem actually establishes a barrier to entry and denies gains from trade in this service.

In summary, the current bilateral system would fail the necessity test as defined here. However, the development of a replacement policy is difficult, and transitional arrangements are complicated given the system currently in place and the interests that have been created around the current arrangements.

D6.2 Options for achieving regulatory reform

Unilateral

The unilateral route is one option, but requires some complementary institution building and cooperation on competition policy. It is interesting to consider the level of international cooperation actually required for unilateral reform air services. The main requirement may only be an efficient domestic regulation system on aircraft slots, for example.

Bilateral

Another option is reform within the bilateral system, especially via the US 'open skies' agreement. Gerry Gentle's paper neatly outlines how this agreement works, although it should have taken a stronger position on these agreements. They are not effective — they are a 'hub and spoke' approach to reform that benefits the hub country like the United States. Yet the dynamics of the process are hard to avoid. If a country signs the agreement, traffic will be diverted to this country, putting pressure on neighbouring countries to also sign the agreement. The US models are quite limited, having no cabotage and avoiding the issues of ownership, which are both critical issues to our trading partners in Asia. The US route is simply about capturing rents more efficiently for US carriers.

Plurilateral

There is further discussion in the paper of the plurilateral route versus the multilateral route. The plurilateral model presented is the open club. The key issue is the accession clause. Will it work? Having set up a club, it is not clear why a group of countries would want to let anybody else in. They could instead exclude applicants and force them to negotiate.

The rules could provide access in the first place, as Gerry Gentle stressed, and if the policy is driven by consumer interest, they could allow new players to enter the market on the same terms. The risk is that producer interests capture the club process. They could define the club membership to suit themselves and implicitly cooperate (in their new alliance structures) to milk the rents even more efficiently.

The fundamental problem with any sort of conditionality that arises in both a bilateral and plurilateral system is that it retains a barrier to entry. One could argue that the plurilateral system is better than the present system, and indeed the Productivity Commission provided model results to demonstrate the gains that are possible, but the models also highlight the conditions necessary to realise these gains. A dynamic process that encourages expanding membership is still needed.

A further question is who writes the plurilateral agreement? Should the foundation members write the agreement or should potential entrants be involved? This is currently being debated in APEC. Providing everyone with the chance to comment on the draft will ultimately lead to a more open outcome.

Multilateral

Finally, what about the multilateral approach? The paper seems to only consider the ‘all or nothing’ outcomes. It would be helpful if the paper provided some intermediate options for making progress in the current GATS review. An UNCTAD meeting last year provided a possible sequence for reform. First, it would be necessary to clarify what activities of air transport were actually covered by the GATS. A better definition might also expand the coverage of the GATS. Second, an effective way to introduce reform may be by moving from the easiest to the hardest forms of delivery; starting with charters, freight, ownership and then moving towards more difficult issues, including delivering from home base. Obviously this would take time, but the outcome may be better than other options.

The dilemma is an interesting one — do you patiently work at the application of multilateral principles or do you permit a plurilateral solution which may be better, but which might also lose momentum and become simply a better way for airline alliances to earn rents? The resolution requires firm action from above and below. It requires progress at a multilateral level through the GATS to review MFN exemptions. It also requires action from below in the form of regulatory review processes of the sort discussed earlier at this conference.

General discussion

The discussion focused on:

- liberalisation under the current bilateral framework;
- liberalisation of ownership and control rules; and
- liberalisation under the GATS.

Liberalisation under the current bilateral framework

One participant noted that while the bilateral system was a 50 year old system devised to deal with situations of the time, the changes that had occurred, and were continuing to occur, were overlooked. It was acknowledged that the current system was not ideal, but that the bilateral system did provide a series of levers with which to create liberalisation.

Gerry Gentle responded by noting that progress was possible in a bilateral framework and, indeed, was preferred to regressing through an illiberal multilateral framework. But a bilateral framework was very slow and highly problematic, and did not tackle fundamental issues of ownership, control and the inability to operate global networks efficiently.

Liberalisation of ownership and control rules

The importance of changing the ownership and control rules to make larger gains in liberalisation was stressed. Once you could buy somebody else's participation in another bilateral market, it became irrelevant if you were excluded under the current regulatory structure. The United States was non-progressive in the context of ownership and control rules.

A view was expressed that future progress was likely to be within a plurilateral framework, driven by the United States and Europe. The transatlantic free aviation area may provide a forum for changing ownership and control rules, though it was noted that, as with the US 'open skies' agreement, it might simply provide the means to collude. Often, the main gain from an 'open skies' agreement with the United States was an ability to gain an antitrust immunity.

Liberalisation under the GATS

There was general agreement that progress under the GATS was likely to be limited in the short term, though the benefits of including air services in the GATS were acknowledged.

It was observed that, by insulating air services from other sectors in the GATS, the ability to make tradeoffs and open the scope of negotiations was lost. All the tradeoffs must therefore occur within airlines and on a bilateral basis between airlines. These tradeoffs were completely excluded from the GATS, with inclusion requiring a favourable vote from two thirds of GATS members.

It was pointed out that liberalisation of tourism services was one of the few issues that had gained consensus within the WTO Council for Trade in Services. This may provide the opportunity to introduce air services negotiations into the GATS framework.

A participant suggested that the GATS might be used to collectively raise ownership thresholds around the world under mode 3, dealing with restrictions on investment. The GATS may provide a vehicle for achieving multilateral investment regime liberalisation in the industry, without having to tackle more complex issues.

PART D

REGULATING FOR
CONSUMER CONFIDENCE

7 Lessons from the Asian Crisis

Jeff Carmichael

Australian Prudential Regulation Authority

7.1 Introduction

The paper is supposed to address the lessons from the Asian crisis for Australia's financial regulatory framework. However, poor regulation actually had little to do with the recent Asian crisis. That does not mean regulatory reforms in the region are not required — indeed, reforms are crucial. There are, however, some lessons for Australia from the episode.

This paper commences with a broader historical perspective about crises. It then turns to the Asian experience, before finally looking at the regulatory lessons.

7.2 Crises — a historical footnote

Superficially, there may be a temptation to surmise that crisis is a peculiar characteristic of the current global financial system. After all, in the past decade alone, there has been the European crisis of 1992-93, the Mexican crisis of 1994-95, and the Asian crisis of 1997-98.

It is very tempting to view the current world financial system as crisis prone and to look to the distinctive features of this period, such as globalisation, deregulation, technology, and so on, as causes of the problem.

While there may be something to learn from changes in the environment, there are much more fundamental issues involved, which are by no means uniquely related to these evolutionary pressures.

Crises, both international and domestic, are by no means a new phenomenon. In a recent paper on this topic, Bordo and Eichengreen (1999) point out that Sprague's (1910) classic study of banking crises in the United States contains separate

chapters on the crisis of 1873, the panic of 1884, the stringency of 1890, the crisis of 1893, and the crisis of 1907.

As seen by Bordo and Eichengreen, the experience of last two centuries can be summarised as follows:

- first, the nature of currency crises has been closely linked to the various international exchange rate regimes that have existed over time; and
- second, banking crises have been largely linked to the credibility of the exchange rate system.

Thus, for example, Bordo and Eichengreen find that:

- currency crises were short and generally mild under the gold standard of the nineteenth century;
- where currency and banking crises coincided under the gold standard, they were invariably severe;
- the inter-war gold standard, which was characterised by declining credibility over convertibility, was notoriously crisis prone, with very severe twin currency and banking crises; and
- under the Bretton Woods system, currency crises were short and mild and banking crises were virtually non-existent.

From this thumbnail sketch, it is important to distinguish between two fundamentally different types of crises:

- a financial crisis involving disruption of financial networks caused by an unstable financial structure — this could arise from a loss of confidence in the exchange rate system or in the banking system or both; and
- an economic crisis directly related to failures in the underlying real economy.

In a perfect world, all crises would be avoided. The reality, however, is that every economy is subject to real shocks of one sort or another that cause economic disruption. It is also reality that all economies sometimes become sufficiently out of balance that a crisis is not only inevitable, it is a necessary cathartic process for restoring balance within the economy.

What really should be avoided is a situation where a financial crisis creates an economic crisis in an otherwise sound economy.

7.3 The Asian crisis — real or financial?

When looking at the recent Asian crisis, the obvious question is what type of crisis was it: a financial crisis that had economic repercussions — or an economic crisis from the start, a crisis that had to happen?

Imagine the following scenario, suitably abridged from Rodrik (1999):

- financial market deregulation sets off a credit boom and an explosion in real estate and equity prices;
- private sector debt rockets up from 85 per cent of GDP to 135 per cent within five years;
- the economy becomes overheated, generating sizable current account deficits;
- unable to borrow long term, the private sector accumulates large quantities of short-term foreign-currency debt;
- real appreciation of the exchange rate weakens exports;
- denying rumours of a depreciation, the government maintains the peg on an increasingly overvalued currency;
- devaluation in a nearby country reveals how vulnerable the economy is to a loss of international confidence;
- suddenly, foreign lenders cut their short-term credit lines;
- the central bank tries to hold the exchange rate, but eventually has to let the currency go, and down it tumbles; and
- the drying up of credit and the currency depreciation trigger a wave of bankruptcies and the economy faces its worst economic crisis this century.

Is this Thailand? Or maybe South Korea? It could be either, but it is not. The country in question is Sweden — and not 100 years ago — this was as recent as 1992-93.

The similarities are nonetheless striking, and serve as a good reminder that financial crises can overwhelm countries that do not suffer from deep-seated structural defects, as well as those that do.

Interestingly, in the post-crisis analysis of the Swedish economy there was no discussion of crony capitalism, corruption, weak laws, lack of transparency and so on. And yet there was no dispute that it had experienced a genuine crisis.

There is also no dispute that unstable configurations, such as that suffered by Sweden and its Scandinavian neighbours, tend to be self-fulfilling. The problem, of

course, is that domestic currency depreciation causes domestic liabilities to increase relative to assets. This, in turn, causes bankruptcies, which feed the pressure for further depreciation.

Drawing this parallel a little further, contrast Sweden in the early 1990s with China over the past two years. A recent study of corruption in China put resource losses from corruption in the billions of US dollars. A study by the Brookings Institution (Lardy 1998) estimates that the last official estimate of the ratio of non-performing loans in the Chinese banking system of over 25 per cent is likely to be biased downwards by a large margin. The same study estimates that the cost of recapitalising China's banking system would run to around 30 per cent of GDP. Few would support the proposition that China has workable business laws or adequate financial regulation (indeed, they are still grappling with the foundational need for accounting laws, bankruptcy laws and a credit culture in their banking system) — and the list goes on (BIS 1999).

Yet during the Asian crisis, China managed to stand firm, not quite unscathed, but with a semblance of stability and certainly avoiding the wild gyrations of its near neighbours.

So what is the lesson here? That good economic policy and management are irrelevant? No, the lesson from the experiences of Sweden, its Scandinavian neighbours, the Asian economies and China is simply that, as long as capital flows remain large relative to the liquid assets held by national governments, and are easily reversible, the economy will be hostage to potentially spectacular financial crises and that these financial crises can lead to economic crises.

Importantly, while poor macroeconomic management may exaggerate this exposure, sound management is no guarantee of immunity.

But was there more to Asia's downturn? With the benefit of hindsight, many have argued that the Asian crisis was a true economic crisis, not just a financial crisis. As supporting evidence they note the extent of corporate and financial failure:

- much of the short-term foreign currency debt was borrowed by banks, many of which have since shut their doors;
- bad corporate debts abound; and
- *ex post* investigation has exposed corruption, poor lending practices, inadequate property and bankruptcy laws, lack of transparency, inadequate accounting standards and poor corporate governance, to name just a few of the underlying deficiencies.

There is no doubt that these weaknesses are all undesirable in terms of economic efficiency, resource allocation and economic welfare. But are they indicative of a crisis that was going to happen anyway? Was there such a deep-seated malaise in these economies that they were fundamentally out of balance and in need of a crisis for the sake of adjustment?

The evidence is unconvincing. It is curious to say the least that many of the proponents of the Asian ‘structural deficiencies’ argument as the key to the crisis were lauding the same countries just prior to the crisis for their outstanding performances. These critics were happy to point out that:

- the fiscal positions of these countries were uniformly sound;
- there was no conspicuous evidence of excess supply of central bank credit;
- economic growth was strong;
- inflation was low;
- unemployment was low; and
- many were liberalising their capital markets in response to prompting from Western economies.

While the jury is still out on the Asian ‘structural deficiencies’ argument, the Asian events bear a much closer resemblance to those in Scandinavia in the early 1990s than to those of Russia a few years later.

The more important question is why financial crises like this occur. Unfortunately, the allotted time does not allow this question to be explored fully. Nevertheless, fixed exchange rate systems have essentially the same fragility that banking systems have — liquidity works well while there is confidence in the ability of the system to perform, but it can evaporate very quickly once that confidence disappears. The parallel between a bank run and a country run is very relevant in the context of the Asian crisis.

The two main lessons for Australia at this macroeconomic level appear to be that having a sound macroeconomy is probably the best insurance against crisis, and that the exchange rate system worked very effectively under pressure as a means of minimising contagion of the Asian crisis to the Australian economy.

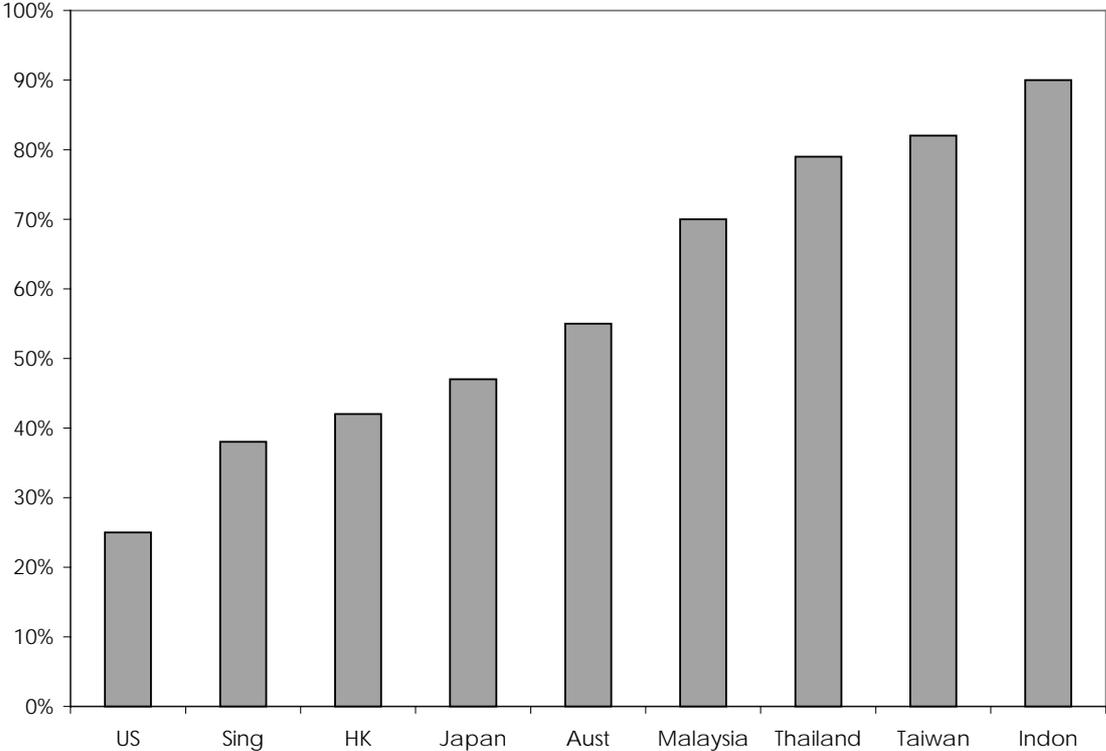
7.4 The role of financial regulation

In moving from the macroeconomic level to the regulatory, the obvious question is: if financial regulation did not cause the Asian crisis, did it play any role at all?

The answer to that is a resounding yes. Once the currency crisis had turned on the heat, the regulatory shortcomings, especially in the banking system, quickly compounded the situation. Indeed, the lack of regulatory credibility further undermined confidence in governments and their exchange rate systems at the very time that they needed the support most.

Before coming to the shortcomings, it is worth noting the importance that the banking system plays in Asia, relative to its role in developed financial systems. Figure 7.1 shows bank assets as a percentage of financial sector assets in selected countries.

Figure 7.1 Bank assets as a percentage of financial sector assets
1998



Data source: Reuters Information Service.

The dominance of banks in Asian markets (with the obvious exceptions of Singapore and Hong Kong, neither of which suffered unduly during the crisis) means that their markets were, and still are, highly vulnerable to problems in the banking sector. As noted in a recent speech by Alan Greenspan to the Financial Markets Conference of the Federal Reserve Bank of Atlanta (1999), the existence of non-bank financial institutions and markets increases financial resilience. In his words:

Multiple alternatives to transform an economy's savings into capital investment act as back-up facilities should the primary form of intermediation fail.

In making his case, Greenspan points to the way in which banks replaced the market in the United States in 1998 and did the reverse a decade earlier. In both cases, the US financial system was able to recover quickly, as alternative means of intermediating savings took over the gap vacated by more traditional means. Greenspan contrasts that experience with the slower response of the predominantly bank-based Japanese economy to its banking problems of the late 1980s and the depth of the more recent crisis in Asia.

Against this backdrop, some of the more striking shortcomings in the Asian banking system that the crisis exposed include:

- inadequate capitalisation of banks;
- excessively large exposures;
- poor corporate governance;
- inadequate covering of foreign exchange exposures;
- poor lending decisions and documentation;
- inadequate collateral;
- inadequate bankruptcy laws; and
- inadequate disclosure.

Most Asian countries had partly liberalised their financial systems in the late 1980s and early 1990s, reducing government controls on capital flows and interest rates, but retaining some government ownership of financial institutions and some control over credit allocation decisions. Several governments continued to encourage banks to lend to favoured projects, industries or companies. In some cases, the liberalisations were poorly sequenced, with prudential controls lagging behind the relaxation of other restrictions.¹ The result was an increase in risk, with big expansions in credit to speculative activities, such as property and equity, without an offsetting increase in capital. Long-term projects were financed with short-term funds, often in foreign currency and risks were not properly assessed.

Interestingly, from a regulatory perspective, many of the problem issues of the period were covered by prudential standards. The real failure appears to have been in the absence of enforcement.

¹ While Thailand, Indonesia and Korea did not, some countries, including the Philippines, Singapore and Malaysia, tightened prudential regulations. For a more detailed coverage of individual country responses during the 1990s, see DFAT (1999).

The absence of enforcement, in turn, was, in a number of cases, a product of the economic environment of the economies involved. A good case in point is Thailand. The banks in Thailand have unusually concentrated ownership by Western standards; in many cases the owners are wealthy families or networks of families. Further, the owners of the banks are also major owners of industrial companies, which are, in turn, the source of large lending exposures for the banks. To compound matters, the same owners play a significant role in the Government.

Resolving problem loans requires a degree of objectivity and a strong enforcement orientation on the part of the regulator. This process is hopelessly compromised when the same parties sit on both sides of the resolution table, as well as playing a role in the discussions for regulatory reform. Although the example applied to Thailand, similar weaknesses apply to several other Asian countries, especially Korea.

7.5 Raising the standard of financial regulation in Asia

Resolving these problems and raising the standard of financial regulation in Asia is a major exercise. In a number of countries the problem has several layers:

- restoring the financial viability of industry;
- building a credit culture within the banking community;
- building the supporting infrastructure of laws and accounting rules; and
- building the regulatory powers and skill base of the regulators.

Regulatory reforms currently in progress focus largely on capital adequacy, limits on foreign currency exposure and accounting for non-performing loans — a relatively narrow focus against the breadth of the problems listed above.

That narrowness notwithstanding, there can be little dispute over the importance of capital. Most Asian countries had implemented the 1988 Basle Capital Adequacy Accord requiring an 8 per cent minimum risk-weighted capital ratio prior to the crisis. However, a number of countries had not enforced the ratio on domestic banks. Further, lax rules for provisioning for non-performing loans meant that reported capital and actual capital was widely divergent in banks that purported to meet the international standard. Korea, Thailand, Indonesia, China and even Japan faced this problem. Most Asian countries are now actively seeking to rebuild bank capital and to tighten the rules on provisioning.

A number of Asian countries are also seeking to reform the powers of their regulators, giving them greater autonomy from government, greater legal powers to

intervene in troubled banks and greater scope to require disclosure and provisioning. Some countries are also exploring the possibility of legislated ‘prompt corrective action’ to tie the hands of the regulators so as to avoid excessive forbearance.

That the resolution of these problems is going to be a long process is underlined by the fact that some of the problems to be resolved, such as tightening the legal and disclosure regimes, are pre-conditions to effective regulation.

It is also worth remembering that the regulatory frameworks that currently exist in developed financial systems such as in Australia were not built overnight. Rather, the regulatory framework emerged as an outcome of several decades of learning from Australia’s mistakes, as well as those of others. This process of learning is still an ongoing one.

7.6 Lessons for Australia

At the risk of sounding dismissive, there are very few direct lessons for Australia from the recent Asian experience.

At the macroeconomic level, the Australian economy proved remarkably resilient and the exchange rate system rode through the turbulence with relative ease. The exposure of Australia’s banking system to Asia was relatively modest and the improvements that had taken place over the previous decade in risk management helped mitigate what might otherwise have been a serious blow to the capital of the Australian banking system. The depth of back-up in the form of non-bank financial institutions was a comfort, although they were not really called on.

Australia’s regulatory system suffers from none of the fundamental weaknesses of those Asian neighbours who felt the weight of the wild swings in currencies and non-performing loans.

There are, however, two indirect lessons for Australia to learn from the recent Asian experience.

The first lesson is that Australia should be prepared to pursue regulatory policies that are unambiguously sensible, even when these meet the criticism that they will make the Australian financial institutions internationally uncompetitive.

It is interesting in this context to recall that one of the driving forces behind the Wallis Inquiry (Wallis et al. 1997) was a concern, raised by industry, that this was indeed happening — that Australian regulation was too heavy handed and that

Australia was losing business to its Asian neighbours who were more ‘flexible’. Not surprisingly, this message softened as the high cost of flexibility became clearer. The problem is that these costs can often take a long time to appear. But the lesson of the Asian experience is that they do appear eventually and often at the very time that the country least wants them to.

The same issue has arisen in the context of the Australian Prudential Regulation Authority’s proposed new standards for general insurers. The proposed standards are designed to bring an outdated Act into the modern world of risk management. A common cry, however, is that Australia has no place trying to lead the world in reforms, however sensible the reforms may be.

In light of the recent Asian experience, Australia should continue to reform regulatory practices, wherever there is a clear need to do so. In the long run, it is better to attract business to Australia because of confidence in the safety, soundness and fairness of Australian markets and institutions, rather than because of any short-term desire to compete on the margin of regulatory laxity.

The second lesson is that, regardless of the strengths of Australian financial institutions and the regulatory system, Australia can never be completely immune from shocks that originate overseas. This is part of the price of being involved in international trade in services.

It is therefore in Australia’s interest to assist, as much as possible, in the rebuilding of the financial and regulatory systems of Australia’s neighbours and to play an active role in discussions of international reform. Not only can Australia help, it has an obligation to help. That help can be as intensive as helping with reforms and building training programs, or as simple as allowing other regulators to send staff to spend time in Australia’s institutions to see first-hand how it is done in other countries.

7.7 Concluding thoughts

Regulatory shortcomings in Asia did not cause the recent Asian crisis.

While they may not have caused the crisis, regulatory shortcomings certainly contributed to the severity of the crisis and have cast a shadow over the speed and reliability of the recovery. In short, without regulatory reforms, many of the Asian economies will remain vulnerable to further financial disruptions.

There are, however, indirect lessons for Australia from the crisis, including the importance of standing ground on sound regulatory reforms, and the need to assist other countries in their reform processes.

While this may sound a little dismissive of the lessons that Australia can learn from the Asian crisis, every crisis contains some surprises and no matter how strong Australians think their financial and regulatory systems are, some improvement is always possible. Perhaps the greatest lesson for regulators to learn is humility.

References

- BIS (Bank for International Settlements) 1999, 'Strengthening the banking system in China: issues and experience', Policy Papers no. 7, Basel, Switzerland.
- Bordo, M. and Eichengreen, B. 1999, 'Is our current international economic environment unusually crisis prone?', in Gruen, D. and Gower, L. (eds), *Capital Flows and the International Financial System*, Reserve Bank of Australia, Sydney.
- DFAT (Department of Foreign Affairs and Trade) 1999, *Asia's Financial Markets: Capitalising on Reform*, East Asia Analytic Unit, DFAT, Canberra.
- Greenspan, A. 1999, Do efficient financial markets mitigate financial crises?, Paper presented at the 1999 Financial Markets Conference of the Federal Reserve Bank of Atlanta, Georgia.
- Lardy, N.R. 1998, *China's Unfinished Economic Revolution*, Brookings Institution Press, Washington DC.
- Rodrik, D. 1999, Governing the global economy: does one architectural style fit all?, Paper prepared for the Brookings Institution Trade Policy Forum, Washington DC.
- Sprague, O.M.W. 1910, *History of Crises Under the National Banking System*, US Government Printing Office, Washington DC.
- Wallis, S., Beerworth, B., Carmichael, J., Harper, I., Nicholls, L. and Smith, G. (Wallis report) 1997, *Financial System Inquiry*, Final Report, AGPS, Canberra.

Discussant — *Tom Valentine*

University of Western Sydney

There is little to disagree with in Jeffrey Carmichael's paper.

A major contributor to the Asian crisis was the attempt to maintain a fixed exchange rate while pursuing inconsistent domestic policies, although it is difficult to convince the countries concerned that this was the case. As the paper noted, Australia's floating exchange rate and responsible domestic policies did much to insulate it from the crisis. Also, the regulation of financial intermediaries in Australia is close to state of the art.

As a result, this discussion focuses on the most pressing problem in the regulation of Australian financial markets, the regulation of investment products.

D7.1 Current problems in regulation of investment products

The prevailing view on these products can be stated briefly. Investment products do not create a risk of systemic instability. In particular, since their prices are market determined, there is no incentive for consumers to stage a 'run' on their providers. The providers are not part of the payments system.

Therefore, the argument for their regulation arises out of asymmetric information — the providers of the products have greater knowledge of their risks than their consumers.

It is becoming clear that the gap between the knowledge of providers and that of consumers is even larger than usually supposed. In part, this is because of the lack of sophistication of most investors. There have been many examples in which investors have lost their entire retirement savings.

The present regulatory regime seems ill-equipped to deal with such problems. This area is largely the responsibility of Australian Securities and Investments Commission (ASIC), and its approaches (and weaknesses) can be summarised briefly as follows.

First, there is a heavy emphasis on disclosure, which may seem appropriate when the problem is viewed as one of asymmetric information. However, the procedures developed by ASIC (such as prospectus approvals and licence applications) put a costly burden on investment promoters without assisting investors. It appears impossible to avoid creative accounting in prospectuses, no matter how elaborate the requirements. The more elaborate prospectuses are, the less likely that the average investor will understand, or even attempt to understand, the information

provided. Elaborate prospectus requirements are one factor retarding the growth of a corporate bond market in Australia.

Second, ASIC appears to have very limited enforcement powers. It is not clear whether this problem arises from weaknesses within ASIC itself, or from its need to rely on the Corporations Law and to operate through the courts. Courts may pay attention to legalistic formalities, rather than to the nature of the underlying transactions. Undoubtedly, ASIC's job is made more difficult by the permissive approach adopted by auditors and by the unjustified weight given to their opinions. Misrepresentations in prospectuses are passed over without penalty. Investors who lose money have little alternative but to pursue long legal actions at their own expense. Such actions are typically unsuccessful, as investors do not have the financial or psychological resources to pursue this arduous path, even when they have legitimate grounds to do so. Licensees retain their licences, even in the face of very questionable behaviour.

Third, ASIC puts considerable emphasis on the education of financial advisers. This is a highly desirable step, but an ability to pass examinations does not ensure ethical behaviour. Effective enforcement is still required.

Fourth, the replacement of the old system of having a trustee for a managed investment by a single responsible entity, which manages the investment, as well as protecting the interests of investors, is fraught with danger. There is a clear potential for a conflict of interest under the new arrangements. For example, responsible entities who are also fund managers will have an incentive to engage in active management, although there is considerable evidence that this reduces the return earned by investors (Drew 2000).

D7.2 Policy options

It is time to review the approach to investor protection. Obviously, this would be a huge task. However, the following possibilities are worth some consideration.

- ASIC should be given increased enforcement powers. At the same time, it should look at simplifying its procedures and shortening the time that they take.
- Investor tribunals could be set up to deal with complaints. These tribunals should be staffed by individuals trained in investment, rather than lawyers, and they, rather than the courts, should deal with cases brought by ASIC. The tribunals should have the power to refer cases to public prosecutors for prosecution as criminal cases.
- An insurance scheme could be established to cover losses arising from fraud by investment managers. To some extent, contributions to this fund could come

from the money saved by not having to meet the present elaborate requirements. In any case, a modest levy would generate the necessary funds. Also, investment managers could be required to post a bond.

Another aspect of the Australian financial system that needs re-examining is the Australian tradition of making lump sum payments to retirees. Many of them do not make sensible investment decisions with those lump sums.

In conclusion, Australia has little to learn from the Asian crisis. Nevertheless, Australia does have an opportunity and a need to make changes to our approach to regulating the investment industry. Australia could possibly lead the world in this area.

Reference

Drew, M.E. 2000, Market efficiency, superannuation and funds management, Thesis, Department of Economics, University of Queensland, unpublished.

General discussion

The discussion focused on:

- lessons from the Asian crisis; and
- regulation of investment products.

Lessons from the Asian crisis

Poor enforcement of regulations

It was suggested that the message from Jeffrey Carmichael's paper was that Australia could actually learn some important lessons from the Asian crisis. The paper pointed out that poor enforcement, not the regulations themselves, contributed to the crisis. A system of prudential requirements that were not enforced created the false expectation that prudential requirements actually existed. Society should be very careful in creating the appearance that regulations were there when, in fact, they were not so. Thus, poor enforcement of regulation, and therefore poor regulation, had a great deal to do with the crisis.

Another participant commented that there might not be a need for the government to act as the regulator, as long as the market was well informed. It was noted that New Zealand had prudential provisions, but regulators did not know how to enforce them. When information about the provisions was made publicly available to banks transacting with one another, this effectively created its own regulatory system, independent of government. Under this system, if a bank thought that another bank was not behaving with sufficient prudence, it adjusted its behaviour accordingly.

Jeffrey Carmichael clarified that the lack of enforcement did not cause the Asian crisis, but certainly contributed to the depth of the crisis and the slow recovery. In relation to the New Zealand system, it was noted that it was easier to leave it to the market to regulate when the country did not own any banks or when the foreign parent companies were regulated in another country. There were also some differences between the theory and the application of these policies, as some banks in New Zealand were unclear of their obligation to provide information to consumers.

Industrial economies

A third participant said that the major concern for industrial countries was the large unexplained decline in output resulting from the Asian crisis. This might have been

because Asian economies had more unhedged foreign borrowing that created problems when the exchange rates declined so much, or because they were unable to restructure their industrial sectors. The participant expressed concern about how well economies such as Australia would handle a big finance-based shock, such as the Asian crisis.

Will Martin suggested that the modelling work he undertook with Warwick McKibbin indicated that the most plausible cause of the Asian crisis was a decline in expectations of future returns, profitability and productivity; in a sense, the end of some sort of irrational exuberance (McKibbin and Martin 1999). If these results were run through the Warwick McKibbin's *G-cubed* model, the decline in investment and consumption and the turnaround in the current account were relatively easy to explain. What seemed harder to explain in transient, sticky wage terms was the big decline in output. They modelled the decline in output as an additional exogenous shock, which reinforced the other effects.

Jeffrey Carmichael responded that the large decline in output had a lot to do with foreign currency liabilities. The fall in the currency and the resulting insolvency affected decision-making at the time. If Australia were exposed to such an extent, then it could have the same problem. Even with a market exchange rate, Australia did have a big currency depreciation, but the resilience of the Australian financial system came from the risk management culture in Australia. Many Australian industries were 100 per cent covered through that period. However, in the absence of figures on the exposure of all Australian industries, this explanation was only suggestive.

Other issues

Some participants noted that the paper indicated that economies would be susceptible to financial crises, as long as capital flows remained large relative to the liquid assets held by national governments. Another participant noted that it was not apparent why liquid assets needed to be held by governments.

Jeffrey Carmichael responded that, in a system where the government held liquid assets to support an exchange regime, it effectively made a promise about convertibility and the credibility of the government was occasionally subject to question. A system promising convertibility needed to have the wherewithal to back it up, otherwise crises would occur. This was not an issue for Australia, as the government did not promise convertibility.

Concern was expressed that figure 7.1 showed Japan to have a lower percentage of banks than Australia, which contrasted with the popular perception that Australia

was not as bank-based as Japan. Jeffrey Carmichael responded that statistics for Japan may have reflected differences in the definitions of the total financial sector or how banks and bank subsidiaries were included in the total financial sector. The statistics would be examined to clarify the issue.

Regulation of investment products

In response to the comment by Tom Valentine, one participant suggested that prospectuses were very useful for exercising a duty of care. In Australia, prospectuses were required to contain *all* information that would be material to an investor in determining whether or not to invest in a company.

The participant also stated that ASIC had enormous enforcement powers. Whether it chose to exercise those powers or not was a matter for ASIC, the authorities and those that dealt with the appointment of ASIC and its funding. Further, the market was already over-regulated in relation to takeovers, mergers, prospectuses and the like, and setting up a new series of tribunals and regulations would create an even more complex position for investors. Ultimately, investors needed to be made aware that investment was a risky business, requiring interest, intellect and wisdom.

Tom Valentine reiterated his view that prospectuses provided little useful information for making investment decisions and that this was an area that should receive attention. He also clarified his comment about building additional mechanisms to regulate investment products. It was more expensive and complex to go through the courts, which may not have the knowledge and experience to make judgments. The process would be more efficient if it were left to those who were trained in the area.

Reference

McKibbin, W. and Martin, W. 1999, 'The East Asian Crisis: investigating causes and policy responses', Policy Research Working Papers, no. 2172, World Bank, Washington DC.

8 Principles for regulating the professions

Neil Byron, Lisa Gropp and Kim Gusberti[†]
Productivity Commission

8.1 Introduction

Rather than simply outlining principles for regulating the professions, an approach that could be construed as presuming that professions should be regulated as a matter of course, this paper steps back and asks the question: do professions need to be regulated at all?

The paper begins by defining what a profession is, and identifying the nature of the regulatory issue. This is followed by a consideration of the non-statutory solutions that can overcome information asymmetries, and issues that must be considered in determining the adequacy of these. The subsequent discussion on options for statutory regulation leads into discussion of the principles. This paper draws extensively on work currently being undertaken for the Productivity Commission's *Review of Legislation Regulating the Architectural Profession*, the draft report of which was released in May this year (PC 2000).

8.2 What is a profession?

The Australian Council of Professions (ACP) defines a profession as:

- a disciplined group of individuals who adhere to high ethical standards and uphold themselves to, and are accepted by, the public as possessing special knowledge and skills in a widely recognised, organised body of learning derived from education and training at a high level, and who are prepared to exercise this knowledge and these skills in the interests of others; and

[†] This paper draws heavily on the Productivity Commission draft report *Review of Legislation Regulating the Architectural Profession* (PC 2000). Additional commentary is the responsibility of the authors.

-
- inherent in this definition is the concept that the responsibility for the welfare, health and safety of the community shall take precedence over other considerations. (ACP 1997, p. 5)

In contrast, the Oxford Concise Dictionary (1968, p. 1593) defines a profession simply as:

... the occupation which one professes to be skilled in and to follow.

... any calling or occupation by which a person habitually earns his living.

Initially, the word profession referred exclusively to doctors, lawyers and members of religious orders. Over time, its use expanded to encompass a variety of providers of expert services and advice. A number of occupations may provide ‘expert advice’, but according to Abbott (1988, p. 8):

People don’t want to call automobile repair a profession because they don’t want to accord it that dignity. This unwillingness probably has less to do with the actual characteristics of automobile repair as an intellectual discipline — which are conceptually quite close to those of medicine — than it does with the status of the work and of those who do it.

Hence, at least in the English speaking world, the term profession has come to represent white-collar workers who have completed ‘long, formal education in a university, providing a rigorous and extended theoretical basis’ (Stevens 1998, p. 26). In addition, Abbott (1988, p. 7) noted that this knowledge is ‘not applied in a purely routine fashion, but required revised application case by case’.

From a policymaking perspective, the relevant characteristic of a profession is not its social status.¹ It is the fact that the provider of the service has specialised knowledge not available to the average consumer (including information about the nature of the service and the provider’s quality and ability), and application of this knowledge involves an element of judgement, particularly in the diagnosis of a problem and the possible remedies.

This potentially high degree of information asymmetry provides one possible rationale for government regulation. Another possible justification for regulation is the externalities that are sometimes generated by the provision of professional services.

¹ Nonetheless, the fact that the ‘professions’ historically were chosen from the higher social classes, or today are accorded such status, is related to the nature of the client–professional relationship, especially the implicit contractual relationship or ‘trust’ underlying it (discussed below).

The focus of this paper is on information asymmetries as the topic of the current session is ‘Regulating for consumer confidence’. Lack of consumer confidence arises when consumers lack sufficient information to make an informed choice.²

Information asymmetries and the professions

In virtually all economic transactions, consumers do not have complete information about the attributes or quality of the good or service they are procuring, and typically have less information than the seller does. Given this asymmetry, there is a possibility that consumers may experience problems because of the incompetence or impropriety of sellers of goods and services (or because of defects in the goods themselves), or through genuine communication breakdowns between the consumers and sellers.

The information asymmetry story is well documented in the literature, so will not be repeated at length here. Suffice to say that the source of the problem is that information is either costly to obtain (perhaps requiring a significant amount of search), or cannot be obtained, given the inability of the buyer to measure *ex ante* the attributes of some goods or services. In the latter case, consumers might only learn about the qualities of the good or service through experience (Nelson 1970). Consumer uncertainty about quality may result in a reduction in both the quality and quantity of goods or services consumed in the market — the well-known ‘lemons’ problem (Akerlof 1970).

In the professions, additional complications are created.

- Consumers might not be able to judge readily the quality of the service, even after the purchase. Problems may only surface over time and be difficult to attribute to the original work, due to the complexity of the relationship between the input and the output.
- The provider of the service is also the provider of information about the consumer’s needs. In this case, the consumer needs to trust the provider.

These characteristics of information asymmetry are referred to as ‘credence attributes’ (Darby and Karni 1973). As Leland noted:

A poor plumbing job might not show up for several years, and there might then be doubt as to whether it was caused by the plumber, by misuse, or by an ‘act of God’.

² In any case, regulating the professions often is an indirect way of addressing externalities. If intervention is deemed necessary on the basis of externalities, standard setting or taxation (for negative externalities), and compulsion or subsidies (for positive externalities), would be more direct approaches than regulating the profession.

The plumbing jobs performed by physicians are presumably even more difficult to assess. (Leland 1979, p. 1330)

Credence attributes are the distinguishing feature of the professions. Consumers rely on professionals to:

- identify the precise nature of a problem (diagnosis);
- determine the best way to deal with it; and
- provide the services needed to solve the problem (Wolfson, Trebilcock and Tuohy 1980, p. 191).

There are a number of opportunities in this relationship for unethical behaviour. The provider may have an incentive to recommend unnecessary services, or provide a lower standard than is optimal (and agreed), particularly where the consequences of poor service may not become apparent for some time. Further, due to the infeasibility of precisely specifying quality, not all aspects of the service can be contracted, so an *implicit* (unwritten) contract exists between provider and client, predicated on trust. Indeed, trust is a defining characteristic of the client–professional relationship. As discussed below, the professions and professional associations have developed as a way to regulate and signal the standards of service providers, where this cannot be specified adequately through the contractual process.

8.3 Assessing the need for statutory regulation of a profession

Although, arguably by definition, there will be information asymmetry in the market for professional services, the need for government regulatory intervention does not immediately follow. Information deficiencies are pervasive, yet most markets continue to function reasonably efficiently. In some cases, the consequences of incomplete information are insignificant. Where the potential harm is significant, there are various means to reduce risks. Rottenberg noted:

It does not appear that competitive markets in the real world serve consumers as badly as the informational asymmetry model suggests they do. Therefore, it does not follow that apparent informational asymmetry should be optimally adjusted for by state enforcement of standards of quality and competence ... (Rottenberg 1980, p. 7)

In assessing the need for government regulation to overcome information deficiencies, there is a need to identify and assess the:

- nature and extent of the asymmetry, including the information search costs of consumers and potential market (non-statutory) solutions;

-
- potential harm which could result from choosing the ‘wrong’ provider; and
 - costs and benefits of potential statutory solutions.

Non-statutory responses to information asymmetries

Information asymmetries can be seen as a form of transaction cost that may inhibit profitable transactions. But given normal profit incentives, secondary markets typically develop to reduce transaction costs and, thus, promote profitable trade. For example, profit-seeking service providers have an incentive to try to maintain or create a market for their service, and there are numerous ways that they will signal their quality to consumers, in order to facilitate consumer search and promote consumer confidence.

Consumer search

Although the qualities of service providers may not be immediately apparent to consumers, they may be able to obtain that information, either through searching or through experience (where there is the potential for repeat purchases).

In most circumstances, consumers will search to some extent. For instance, consumers seeking an architect or other building design professional look at a potential provider’s previous work (perhaps through site inspection, looking at publications that feature building design work, etc), or may seek information from industry bodies such as the Royal Australian Institute of Architects. The ability and willingness of consumers to search depends on both the costs and benefits of doing so. The perceived benefits of searching (and, hence, the willingness of consumers to undertake it) will be related to the nature of the service and the financial outlay (size of the investment). The amount of search is likely to be greater where:

- the service is a non-standard one, that is, consumers desire a differentiated service; and/or
- a substantial financial outlay is involved.

For instance, consumers seeking a building design professional desire a differentiated and personalised service. Hence, consumers are likely to search to find a professional whose design style best meets their particular tastes and needs. In addition, the normally large size of the investment for most consumers means that there is a strong incentive to search.

The ability of consumers to search is related to the costs of doing so, and the time available to make a decision. Costs will be high where there is a large variance in

the quality of service providers, a large degree of skill is required to evaluate the service, and there is little ongoing contact between service provider and consumer (Moore 1961). The time available to make a decision depends on the urgency of the service. For instance, the decision to procure building design services is generally not an urgent one. In contrast, a person requiring emergency medical attention does not have time to make a considered choice as to the provider of the service.

The nature of the service and the search undertaken influence the optimal way that information should be provided: either by the market or government, and, for instance, whether information is best provided through generic or differentiated labels.

Warranties or guarantees, and contingent contracts

Warranties or guarantees can be offered by sellers to signal the quality of their products or services, as one way of overcoming information asymmetries. They can be useful when quality is not easily assessed before purchase (because, *ex ante*, any statements are not easily verifiable), but there are measurable performance standards that can be assessed after the service has been provided. To be feasible, then, the output generated by the service provider must be able to be assessed objectively.

Contingent contracts specify payment conditional on the quality of the service provided. As with warranties, they can be effective only where objective standards can be applied. Penalty clauses can be seen as a form of contingent contract where a lower fee is paid if deadlines are not met. In the construction industry, such contracts help to alleviate the problem of time overruns (since completion time is measurable), but it would be infeasible to enforce such a contract for the ‘aesthetics’ of design.

Independent certification agencies

Independent certification agencies can endorse certain service providers, signalling that the provider is of a certain minimum standard. Providers of a range of services signal their quality and standards of behaviour in this way. For instance, car insurance providers maintain lists of ‘approved car repairers’, which can be accessed by consumers. For this to be an effective mechanism to promote consumer confidence, the endorsing agency itself must be considered by consumers to be reputable.

In the Australian building and construction industry, a private body, the Building Surveyors and Allied Professionals Accreditation Board Inc. (BSAP), is recognised

in some Australian jurisdictions as an accreditation agency for building surveyors.³ BSAP also offers an accreditation service to other professional associations within the industry. The incentive to set appropriate standards is high, since the accreditation agency shares legal responsibility if problems arise with an accredited provider.

Reputation

Providers may develop a reputation for good (or bad) service provision, either in terms of their output or the manner in which they provide the service. This can develop either through the direct experience of consumers (where repeat purchases are likely) or through word-of-mouth (recommendations of family and friends). The consequences of a bad reputation for future business may constitute sufficient incentive for ‘professional conduct’, assuming the provider is ‘in for the long haul’.

One of the advantages of the reputation mechanism is that, for many services, it can work where contingent contracts cannot, because ‘reputations can be based on information that does not constitute evidence for the purposes of enforcing contracts’ (Shapiro 1986, p. 845). Furthermore, reputations can work where guarantees cannot be provided because objective standards cannot be applied to a service (it is not clear, for instance, how one could guarantee something as subjective as ‘good’ design).

Large firms and brand names also may act as signals of reliability, by indicating the seller’s ability to maintain long-term customer relationships. Brand names also act as a way of differentiating the service of providers. Although brand names may develop even for standardised goods (services), they tend to be used more where fewer aspects of a good (service) are suited to standardised measurement (Barzel 1982).

Advertising often is restricted in the professions — whether they are self-regulating or regulated by government — suppressing the development of brands at an individual firm or practitioner level. The use of branding, then, generally is restricted to developing a ‘brand’ for the profession as a whole, or for specialisations within the profession. When the profession’s services are essentially homogeneous, and consumers value relatively uniform quality in providers, generic branding may be efficient because it reduces the need for individual search by consumers (Barzel 1982). However, where services are innately heterogeneous (eg for building design), generic branding may offer little assistance to consumers, who

³ The functions and accreditation services of the BSAP Board are due to be taken over by Development Industry Accreditation Services Ltd (DIAS) in 2000.

generally will have to search further, subsuming the information provided by the generic label.

Professional associations and self-regulation

As already noted, the emergence of professions and ‘professional’ conduct can be regarded as a response to information asymmetry and, in particular, the credence qualities of the services provided by the professions.⁴

Membership of a professional organisation with a code of conduct or ethics is one way of signalling to the consumer the reliability, ethics and standards of the service provider. The code may provide some assurance that the professional will hold the client’s interest above any opportunity to achieve unscrupulous financial gain and that the professional will be competent and diligent in providing the service.

Although professional associations can (and often do) coexist with government regulation, the professional association also can form the basis of a self-regulation model. The Council of Australian Governments noted that, in some cases, codes of conduct may be ‘useful *alternatives* to statutory regulation, and transfer the administrative costs from the public sector regulators to industry participants’ (COAG 1997, p. 62, emphasis added).

In a self-regulation framework, professional association(s) are involved not only in setting and maintaining the standards of their members through codes of conduct, but also in devising methods of differentiating the service of their members from those of other providers and/or associations — a ‘labelling’ function. This may be done in conjunction with an independent accreditation agency (as in the example of surveyors discussed above).

In many ways, this is similar to the role that could be played by a government regulatory agency. However, in the absence of statutory licensing (which restricts professional practice), professional associations do not have a captive membership market: membership is voluntary and (most important in distinguishing this from government regulation) any particular association can face competition from other associations representing the profession. Thus, there are incentives to:

- provide informative label(s) to differentiate the profession and members of particular associations, and provide ‘value for money’ services to members; and

⁴ Indeed, the fact that ‘professionals’ initially were drawn from the higher social classes was one way of signalling to consumers that the service providers were above abusing the trust of clients for personal gain.

-
- maintain the credibility of the label in the marketplace by enforcing appropriate standards of behaviour and conduct (this would be done through controls on admission to the association and disciplinary measures).

Self-regulating professions also could be involved in accrediting courses and operating professional development programs, as is done by, for example, accountants and engineers. All self-regulatory arrangements are subject to general consumer protection legislation, such as the Trade Practices and Fair Trading Acts.

The main advantage of self-regulation is contestability and the resulting incentive for the profession to respond quickly to changes in the broader environment — for example, changes in consumer requirements or the emergence of competitors in the industry.

Potential problems have also been raised in the literature. The self-regulating profession will set minimum entry standards (such as educational requirements), in much the same way as government regulation. Shaked and Sutton (1981) noted that, if the profession aims to maintain ‘high’ income levels for its members, there is a potential for these standards to be set unnecessarily high, decreasing the size of the profession below the social optimum. However, unless government regulation restricts practice to those ‘certified’ by the association, ‘para-professions’ can develop independently of the profession, mitigating the potential for monopoly rents to be earned by the latter.

It is also argued that there could be a perception among consumers that a self-regulating profession operates in its own, rather than the public, interest (particularly when it comes to administering disciplinary proceedings). However, the opportunity for capture under government regulation (discussed below) means government regulation is not immune from this problem, and, indeed, may be more susceptible to it where there is no contestability in accreditation. Research undertaken for the Commission’s review of architects legislation suggests that many self-regulating professions, such as engineers and accountants, appear to conduct more visible, transparent and effective quality control mechanisms than some of those regulated by statute.

Although non-statutory solutions alleviate information asymmetries, they will not remove them entirely — indeed it may not be optimal to eliminate them. Crucial to determining the community’s tolerance for information asymmetries (and, hence, assessing the sufficiency of market mechanisms or the preferred type of government regulation) is the potential harm that could result from a decision based on incomplete information.

The potential harm resulting from asymmetry

The types of issues or questions that arise in determining the potential harm fall into three categories.

- *What are the possible consequences?* The consequences of a poorly informed choice of service provider may be relatively minor (eg a small financial loss suffered when an accountant does not inform clients of all tax deductions they are eligible for), or severe (eg the choice of an incompetent surgeon could be life threatening), and may affect a large number of people who are not parties to the initial transaction (ie externalities may be involved).
- *What is the likelihood of the harm occurring?* Consideration of this question can help assess the expected costs of the harm. The probability of harm occurring reflects the degree of discretion enjoyed by the practitioner and the scope of practice (Graddy 1991, p. 689). Any general laws affecting the profession or the industry in which it operates — that is, those not specifically targeted at the profession — also may affect the likelihood of harm.
- *Can the harm be remedied appropriately and effectively?* For instance, a leaking roof can be repaired, while financial redress for many types of loss can be sought through general consumer protection legislation or common law, etc. However, where death is a potential outcome of services provided by an incompetent or unqualified provider, financial redress may not be appropriate compensation for the type of loss suffered. In this case, solutions that try to prevent problems occurring might be preferred to those which simply provide for redress if and when a problem arises.

The importance of identifying the source of potential harm

The professional often provides a range of services, each of which may have its own specific possible consequences, likelihood of harm and adequacy of remedies. Further, in some industries, a number of service providers combine to provide the service ‘package’. In order to determine an appropriate regulatory response for a particular profession, the specific source of the potential harm needs to be isolated. This involves identifying:

- the range of possible services offered by those practising the profession (eg lawyers may offer conveyancing services, mediation in family law, or defence in criminal cases; while doctors (General Practitioners) can administer injections or perform moderately complicated medical procedures); and/or
- the types of possible risk or loss that could be suffered through the substandard provision of the service, and the extent to which the profession in question is responsible for, or has control over, these.

This helps to clarify exactly where non-statutory responses may be inadequate, and enables government to target the statutory response to the ‘real’ problem at hand. Where members of a profession provide a number of services, it may be optimal to adopt different approaches for the different functions. For example, it may be better to regulate relatively routine services by function (eg because non-professionals can also provide the service), and regulate inputs where a greater degree of discretion (and, hence, human capital) is required by the service provider.

The Commission’s draft report on the review of legislation regulating architects (PC 2000) provides an example of how identifying the types of risk or loss helps to clarify the potential benefits of (and need for) regulation. Three sources of risk in the provision of building design and related services were identified: financial, health and safety, and design (aesthetics and functionality). The report suggested that the potential harm is more significant in the area of health and safety, yet this is not the area for which architects typically have sole or ultimate responsibility. Moreover, objective performance standards for building health and safety are feasible and more directly targeted by other instruments, such as building codes. Architects’ area of expertise and comparative advantage is in design aesthetics and functionality, where the potential harm is likely to be far lower (especially given the opportunity and incentive for consumer search due to the nature of the service), and where regulation can do the least, given the innately subjective nature of design.

8.4 Forms of government regulation

The preceding discussion has emphasised the need to consider systematically whether government statutory intervention might be warranted for a particular profession, and if so, in what form. In broad terms, intervention might be justified where the information asymmetry cannot adequately be remedied by non-statutory mechanisms, and the potential harm is significant.

The main statutory options are discussed in this section. They include regulation of performance (which focuses on the outputs produced by the profession), and regulation of inputs (regulating the profession itself, through mechanisms such as registration, certification and licensing).⁵ In addition, generic consumer protection legislation may apply to the profession. The distinctions made between registration, certification and licensing are based on their potential effects on the supply of service providers and, hence, on competition. This differs from the literature (and

⁵ Wolfson, Trebilcock and Tuohy (1980) provides an overview of regulatory options for the professions.

some general usage) where registration and certification tend to be used interchangeably.

Generic consumer protection legislation

Government may choose to provide generic consumer protection legislation, such as Trade Practices and Fair Trading Acts, which can apply to all professions, rather than enact legislation specific to a particular profession. Relying on general legislation may have some disadvantages, including that regulators may not have specialised knowledge in the relevant area. However, there may also be some advantages, including the fact that it is less likely to be subject to capture by the profession and tends to be less prescriptive than profession-specific legislation.

Enforcing quality standards on outputs

Setting and enforcing quality standards on output through codes and legislation is one way of trying to address the issue of quality in markets for professional services. Processes that could be established to support these standards include independent inspection and audit mechanisms (Wolfson, Trebilcock and Tuohy 1980).

This form of intervention has the advantage of focusing directly on outcomes, in contrast to the options for regulating inputs (discussed below) which, to be effective, require very strong links between input and output. Such standard setting could be particularly important in the context of changing technologies in an industry, where the role of particular inputs (and professions) changes, but certain objective output standards need to be maintained. For instance, the role of architects in the building industry appears to have changed over time: where once architects were the main project managers, often they are now one of many players on projects, with others having ultimate responsibility for health and safety standards. In this case, regulating standards for the structural aspects of construction, through building codes and the like, may be preferable to regulating an ‘input’, the architect, who has only partial responsibility, at best, for the source of potential harm.

However, enforcement and monitoring costs can be high and it may be difficult to set standards at appropriate levels. Given this, quality standards might be best used where the product or service provided by the professional is amenable to the setting of objective, measurable standards (for instance, structural safety requirements for constructions), the potential harm is high (perhaps involving externalities), and there is an indirect or imprecise link between the input and the output. In many

cases, the outputs provided by the professions are not amenable to the setting of measurable standards, hence, the more common use of input regulation.

Registration

Registration involves the maintenance of a list (a central registry) of those engaging in and ‘qualified’ to provide a particular service (Friedman 1962). Registration usually is distinguished from certification and licensing by the fact that standards required for eligibility tend to be much lower (perhaps nothing more than having a registered business number). In addition, registration would not necessarily involve a restriction on the use of title. Fees are paid to maintain registration.

Essentially, registration plays an information role: informing consumers of the existence of providers (so providing a screening device and reducing search costs), and providing contact details should anything go wrong. Given the low requirements for registration, though, it might only be an appropriate approach where there is a potential for information asymmetry but the harm is not particularly high.

Certification

Certification involves the reservation of a title to people who have met certain educational and training requirements, considered to indicate a minimum level of competence. Others can legally compete with the ‘certified’ group, but must not use their title. In practice, the requirements for certification tend to include formal education, as well as postgraduate training and experience. There also may be ongoing requirements, such as adherence to codes of conduct or compulsory professional development.

Suggested advantages of certification include the following.

- It can provide some means for consumers to distinguish among groups of providers, while still leaving them the ultimate choice about desired quality. Hence, certification could decrease consumer search costs, particularly where a significant degree of search would otherwise be required to obtain information about the qualifications, expertise, etc of a provider. However, to be effective in promoting consumer confidence, consumers need to know exactly what information the title conveys.
- It does not preclude the development of other organisations representing ‘uncertified’ providers or ‘para-professions’, thereby reducing (but not necessarily eliminating) any anticompetitive effects.

-
- A government certification system also may have more credibility in the eyes of consumers than one maintained by a professional association (though, as discussed below, this is not guaranteed).

Statutory certification also has a number of potential weaknesses.

- It cannot protect consumers from allegedly incompetent providers, because consumers can choose to hire an uncertified provider. If the potential harm from a poor choice is significant, certification may be inadequate.
- It effectively implies that there are discrete classes of provider (ie there is no 'overlap' in their quality) when, in reality, there may be a continuum of quality, and some overlap across categories of providers. Thus, it may not accurately inform consumers of differences in quality among providers.
- It creates a generic title for a group of providers, which may not be appropriate for an innately heterogeneous service. For instance, the services provided by architects are not homogeneous, so consumers normally would still undertake additional search to find the provider best suited to their particular needs. The information sought through consumer search may subsume the generic information provided by certification.
- For certification to be effective, there must be a direct link between the profession being regulated and the harm being targeted. Otherwise, although certification might maintain the professional standard and competence of practitioners, this does not mean that those standards translate into a reduction in potential harm.
- Certification requirements must directly improve the ability of the profession to address the harm being targeted. Where there are no ongoing requirements for periodic re-certification, this link may be tenuous.
- Statutory certification may be used as a *de facto* vehicle for restricting practice.

Many of the perceived advantages of certification can be achieved through non-statutory means, without necessarily suffering from the same weaknesses as statutory certification. The question may be whether such a system can be as credible as a government-backed labelling system.

Licensing

Licensing restricts the performance of certain tasks (practice) to those with certain qualifications recognised by government. Thus, unlike certification, licensing restricts consumer choice to a provider deemed to be qualified.

Potentially, licensing can play a more significant consumer protection role than certification. However, licensing, by restricting entry, inevitably decreases competition in the market (and is, therefore, likely to increase prices to consumers). It also may be used as a vehicle by the profession to raise entry barriers beyond those needed to benefit the community. In addition, as with certification, any potential benefit of licensing is diminished (and costs to the community increased), where there is not a direct link between the harm being targeted and the services provided by the profession, or where qualification requirements do not improve directly the ability of the profession to address the potential harm.

Given the restrictive nature of licensing, it should only be used under very specific circumstances:

- where there are high information asymmetries (so that there is a high probability of a ‘poor choice’ being made — perhaps because consumers tend to underestimate the probability of harm from using unqualified providers (Graddy 1991, p. 678));
- consumers do not have much time to make purchase decisions; and
- the potential harm is high.

To minimise anticompetitive effects, licensing could be targeted at particular services provided by a profession, where the potential harm is particularly significant (rather than the whole range of their services).

Licensing can operate in a so-called co-regulation framework. In this case, government delegates the responsibility for accreditation, and the setting of entry and conduct standards to the profession, and establishes regulation restricting practice to those deemed suitably qualified by the profession. For instance, the profession determines entry standards for doctors, and government restricts the practice of medicine to those people. This approach might be justified where the potential harm is high, but the highly technical and specialised nature of the discipline means that the profession is best able to determine what is required to be a ‘qualified’ provider. The risk, of course, is that, without effective independent monitoring, the profession will act as a monopolist and restrict entry beyond the level required to maintain reasonable quality standards.

8.5 Principles for regulating the professions — characteristics of good regulation

On balance, it might be decided that non-statutory responses are inadequate to address the problems identified, justifying regulation of the profession by government.

However, although statutory solutions might have perceived advantages in promoting consumer confidence (including credibility and perceived impartiality), they also can have a number of potential problems. Credibility is not guaranteed when the regulator is subject to capture by the profession, there is no provision for contestable accreditation, and procedures are not independently or transparently administered. Government regulation also can be inflexible, which is particularly problematic in industries that are changing rapidly. Thus government regulation may inhibit innovation. Further, creation of a legislated monopoly (even if only over the use of a title) can create a sense of complacency in a profession, making it less responsive to consumer needs.

These potential problems with government regulation form the basis of the principles for good regulation discussed below. The discussion incorporates examples from the Architects Acts currently being reviewed by the Productivity Commission. In a sense, these principles are really no different from those for any regulation — all that differs is the context.

Targeted

Good regulation (in general) must be targeted to the problem at hand. To do this, the objectives of legislation must be specified clearly at the outset. This is where it becomes crucial to identify the potential harm and unbundle the functions performed by the profession. Given the difficulty in establishing direct links between the profession and desired outcomes in some cases, setting standards on outputs may be preferred to regulating an input (a member of a profession), when the potential harm is high. For instance, building codes may provide a better (more direct) approach for addressing health and safety issues than regulating a profession (such as architects), which, at best, is only partially responsible for meeting those standards.

Targeting the potential harm may mean regulating some services by function, rather than regulating the profession that provides them. Importantly, ensuring legislation is targeted also helps to minimise any anti-competitive effects, by limiting regulation to areas where it is of most need and most effective.

Independence, transparency and accountability

Credibility is claimed to be one of the main advantages of government regulation, yet credibility can be quickly diminished by inappropriate design of the administering system. For instance, granting a statutory monopoly to a certification body may not generate sufficient incentives for the public interest to be promoted above the profession's interests. *Perceptions* of independence, transparency and accountability are crucial in maintaining consumer confidence. *Actual* independence, transparency and accountability are critical for maintaining professional standards of conduct — if it is expected that poor performance will be detected and punished, then a professional is more likely to adhere to standards of conduct.

There is no guarantee that either perceived or actual independence, transparency and accountability is achieved by government regulation. Registration and disciplinary boards often are composed predominantly of members of the regulated profession. Although the presence of members of the profession on the relevant boards can be justified because of the technical nature of accreditation, certification and complaints processes, the potential for conflicts of interest arises where there is no contestability in accreditation.

Hence, it is crucial to impose measures to ensure the independence, transparency and accountability of the system. This may include having consumer and other non-profession representatives on boards, as was done with the Architects Boards in the United Kingdom. (An issue that needs to be overcome with this approach is what incentive consumers have to be on the Boards — this incentive may not be as high as for members of the regulated profession.)

There are a number of other ways to increase the independence, transparency and accountability of the system.

- There should be clear and specific grounds for disciplining a professional, including incompetence and negligence.
- Procedures for consumer complaints should be widely publicised and should not be unnecessarily complex.
- Reasons for decisions should be made clear and public, while there should be independent avenues of appeal — for client and service provider.
- It should be subject to ongoing external review to demonstrate that the benefits continue to outweigh the costs.

Flexibility

Government regulation is formed at a particular point in time — to address particular issues in a particular social and regulatory context. It does not generally have the flexibility of non-statutory solutions to adjust to changes in the broader environment — and it is difficult to establish a framework incorporating such flexibility. Over time, a lack of flexibility may make regulation irrelevant at best or harmful at worst (eg restricting entry and choice where it is no longer necessary). Regular external reviews may help diminish inflexibility.

Improvements in technology have allowed easier access to information so that non-statutory solutions may have become more effective in many cases. In addition, changes in the industry can mean other specialised occupations develop, changing the role of the regulated profession. Changes in the regulatory context may mean that other, more recent, legislation could better promote regulatory objectives. As discussed above, the role of architects on building projects appears to have changed since the Architects Acts were introduced. In addition, when these Acts were formulated, protection through fair trading and trade practices legislation, building codes and planning laws, either did not exist or was considerably weaker than now.

Visibility (consumer awareness of the system)

If one role of the regulatory system is information provision, then the system fails if consumers do not know exactly what information it conveys. From a consumer protection perspective, the precise detail of a quality enhancement system may not need to be known for it to provide protection or maintain consumer confidence and professional standards of conduct. However, consumers must know of the system's existence and understand broadly the degree of protection offered. To promote credibility, consumers must know the body to which complaints must be addressed. For instance, the existence and role of the Architects Boards in most states does not seem to be well known to the consumers they are intended to inform and protect, so a number of consumers seem to be using other general consumer complaints mechanisms.

If consumers are aware of the system but do not fully understand the provisions of the regulation and functions of the administering (disciplinary) bodies, they may perceive incorrectly that the regulation guarantees the standards of any professional whom they may choose to employ. As noted above, regulation such as registration and certification does not guarantee quality, particularly where the link between initial input (registration or certification requirements) and lifetime output is indirect. If consumers misconstrue the information supplied, they may decrease

their own search effort. Thus, a poorly understood or dysfunctional regulatory system may be hazardous to consumer welfare.

Cost-effectiveness

It goes without saying that professional regulation should be cost-effective; this principle underlies the whole framework for considering the necessity for regulation, as well as developing the regulation itself. And, of course, costs and benefits must be broadly considered. Costs include not only the financial and resource costs of administering the system, but also the compliance costs of the regulated profession and the costs generated by possible negative effects on competition and innovation. For instance, ownership restrictions on architectural practices may have limited the flexibility of architects to organise the way they deliver their services. In addition, inconsistency in ownership provisions across jurisdictions has imposed costs on those attempting to register in more than one jurisdiction.

International consistency of regulation

International trade in services appears to be increasing — with provision occurring either electronically or through the movement of people (who buy or provide the service). In this setting, an important issue is whether Australia should adopt the same standards as other countries so as not to jeopardise the exports of a particular profession's services. However, the cost of any regulation on the domestic economy must be weighed against the benefits derived from promoting exports, as well as other benefits. Further, it is not always necessary to impose the same standards across all countries: standards for goods differ across countries, reflecting local requirements, as do standards for many professional services. Although mutual recognition agreements between countries may be more difficult to negotiate when regulatory standards are different, successful negotiation is feasible. For example, engineers are subject to different forms of regulation throughout the world, yet they have established several unilateral and multilateral trade agreements. There are a number of ways of promoting exports by reducing barriers to trade, through the General Agreement on Trade in Services (GATS) or through regional associations, for example.

8.6 Conclusion

There can be no presumption that all professions need to be regulated by government.

Before considering legislation, it is necessary to investigate what the market can do — identifying when non-statutory solutions are feasible and when they are not, bearing in mind that the very existence of so-called professions is a market response to information problems. Although information gaps will remain, government intervention — which itself can only partially plug the gaps — is warranted only if significant harm could result, and if government can address problems effectively. In addition, the sources of risk should be specified, enabling identification of where legislation could be of use. Looking at the regulatory options available then allows analysis of how and if regulation of the profession can help achieve regulatory objectives.

Where there may be a case for government intervention, the net benefits are likely to be higher where regulation is:

- targeted to the potential harm that it is trying to minimise;
- independently and transparently administered, and subject to external review;
- visible — that is, known and understood by consumers;
- accurate — in terms of the information provided;
- flexible — in the sense of not discouraging innovation;
- designed to minimise any anti-competitive effects — leaving the stronger, more restrictive, regulation for areas where the potential harm is greatest; and ultimately
- cost-effective.

Every option — statutory and non-statutory — has its strengths and weaknesses, benefits and costs. The approach adopted should maximise the net benefits, bearing in mind the particular circumstances of each case.

References

- Abbott, A. 1988, *The System of Professions: An Essay on the Division of Expert Labor*, University of Chicago Press, Chicago and London.
- ACP (Australian Council of Professions) 1997, 'Can the professions survive under a National Competition Policy? The Australian Council of Professions' view', in ACP, *National Competition Policy and the Professions*, Canberra, pp. 1–7.
- Akerlof, G.A. 1970, 'The market for "lemons": quality uncertainty and the market mechanism', *Quarterly Journal of Economics*, vol. 84, no. 3, pp. 488–500.
- Barzel, Y. 1982, 'Measurement cost and the organization of markets', *Journal of Law and Economics*, vol. 25, no. 1, pp. 27–48.

-
- COAG (Council of Australian Governments) 1997, *Principles and Guidelines for National Standard Setting and Regulatory Action by Ministerial Councils and Standard-Setting Bodies*, November (amended).
- Darby, M.R. and Karni, E. 1973, 'Free competition and the optimal amount of fraud', *Journal of Law and Economics*, vol. 16, no. 1, pp. 67–88.
- Friedman, M. 1962, *Capitalism and Freedom*, University of Chicago Press, Chicago.
- Graddy, E. 1991, 'Toward a general theory of occupational regulation', *Social Science Quarterly*, vol. 71, no. 4, pp. 676–95.
- Leland, H.E. 1979, 'Quacks, lemons, and licensing: a theory of minimum quality standards', *Journal of Political Economy*, vol. 87, no. 6, pp. 1328–46.
- Moore, T.G. 1961, 'The purpose of licensing', *Journal of Law and Economics*, vol. 4, pp. 93–117.
- Nelson, P. 1970, 'Information and consumer behaviour', *Journal of Political Economy*, vol. 78, no. 2, pp. 311–29.
- PC (Productivity Commission) 2000, *Review of Legislation Regulating the Architectural Profession*, Draft Report, Melbourne.
- Rottenberg, S. 1980, 'Introduction', in Rottenberg, S. (ed.), *Occupational Licensure and Regulation*, American Institute for Public Policy Research, Washington, DC, pp. 1–10.
- Shaked, A. and Sutton, J. 1981, 'The self-regulating profession', *Review of Economic Studies*, vol. 48, no. 2, pp. 217–34.
- Shapiro, C. 1986, 'Investment, moral hazard, and occupational licensing', *Review of Economic Studies*, vol. 53, no. 5, pp. 843–62.
- Stevens, G. 1998, *The Favoured Circle: The Social Foundations of Architectural Distinction*, MIT Press, Cambridge, Massachusetts and London, England.
- Wolfson, A.D., Trebilcock, M.J. and Tuohy, C.J. 1980, 'Regulating the professions: a theoretical framework', in Rottenberg, S. (ed.) *Occupational Licensure and Regulation*, American Institute for Public Policy Research, Washington, DC, pp. 180–214.

Discussant — *Jeffrey Keddie*

University of Melbourne

There is little in Neil Byron's paper to disagree with, except, perhaps, his overwhelmingly fair portrayal of the professions. These comments reflect my previous experience as the registrar of the Victorian Architects Registration Board.

D8.1 Asking the right questions

Asking the right questions is crucial for professional regulation. Nothing can be assumed. Asking 'do professions need to be regulated by the government?' presents a very unwelcome challenge. It makes for uncomfortable discussion with regulated professionals and those who accept the case for regulation, as if one questioned the Copernican system. They do not, by and large, rely on arguments about information asymmetries, externalities, spillovers and public goods, but rather on history and a sense of injustice when entitlements are threatened. There is no widespread understanding or acceptance that regulation must be justified within a context of National Competition Policy. 'We', they say, 'were here first'.

D8.2 What do definitions do?

Another issue to consider is the question of definition. Dictionary definitions describe, but do not prescribe — that is the overwhelming trend in modern lexicography, despite the popular use of dictionaries to 'settle' questions of meaning. The law, however, does use definitions prescriptively, and the definition can be highly artificial.

In this connection, the distinction between 'registration', 'certification' and 'licensing' is questionable. This distinction introduces another level of information that must be understood by consumers. Introducing more terms means that there will be more consumers being confused by this information.

On the question 'is a register evidence of professional standing and competence?', my experience as a registrar dictates that this was exactly what the register of architects was, because the Act said so. If it is felt someone is incompetent or unprofessional, regulators are supposed to inquire into their conduct and strike them off the register. Of course, this is not always done, but that is what Acts say, so regulators are supposed to do that.

D8.3 What does the State do?

Questions of rationale and definition bring us to the government role. Schattschneider (1960, p. 140) said:

The most important thing about any democratic regime is the *way* in which it *uses* and exploits popular sovereignty, what questions it refers to the public for decision or guidance, how it refers them to the public, how the alternatives are defined and how it respects the limitations of the public. A good democratic system protects the public against the demand that it do impossible things. The unforgivable sin of democratic politics is to dissipate the power of the public by putting it to trivial uses.

This can be applied to regulating professions.

Outside of law and lexicography, definitions have another role in the professions. In Acts for professions, definitions create a territory or space. Put simply, and as a hypothesis, legal definitions can create identities — an architect, a psychologist, a medical practitioner — and so create a territory from which competitors are excluded. Definitions also confer privilege in the discourse, reserving ‘space’ in a field of activity *and* in the language. In a commercial sense, registration provides ‘branding’. This raises the question ‘is this the business of the government’ and, if it is, ‘what are the grounds for government involvement’?

The definition of a profession is not historically fixed. There are values embedded in reliance on criteria such as ‘long formal education in a university’ (Stevens 1998). This raises other questions — how long is long? And what is a university? If the knowledge is ‘not applied in a purely routine fashion’ (Abbott 1988), where does this leave many medical practitioners, architects and teachers, who proceed on a routine basis for much or all of their time. Is there a spectrum to being a professional, according to the degree of ‘routineness’ of activity?

Much regulation of professions is ‘historical’ in the sense that it begins at particular times in response to particular circumstances. Notably this is when there is a significant pressure group favouring regulation — usually representing the profession, rather than the public — and a receptive government, which, if not too enthusiastic, at least sees no reason to resist. The establishment of professions is ‘establishment’ in another sense: it is ultimately related to status. Witness the number of ‘institutes’ granted ‘Royal’ as a title prefix to accompany a charter. In Australia, these marks of ‘specialness’ are part of the symbolism of culture and civilisation. They mirror British institutions (eg the Royal Institute of British Architects and the Royal Australian Institute of Architects), and declare participation in the empire. They also imply recognition at ‘source’ and confer

recognition locally through this connection. They recreate understood forms and symbols.

D8.4 Information asymmetry

The information asymmetry argument presents problems, because it is unclear where the lines are drawn. When is an ‘expert’ a technician and when a professional? To come at the problem from a slightly different direction, why does society have little difficulty in recognising an engineer, down to sub-specialisations, without the aid of regulation? Is it because someone else, the government, the employer, who are presumably more knowledgeable, judge this to be so? This merely shifts the problem along the line. Perhaps it is the luck of the draw — there are many engineers and they have a high-profile professional association.

Do consumers use the information made available by the registration system to address their problems? Experience indicates that they simply do not want to know. If a (former) registrar of architects tells a consumer that X is an architect and that it would be sensible to check X’s previous work *and* previous clients’ views, some consumers do not want to know about problems. If only part of the information — that is ‘X is registered, end of discussion’ — is provided, this may be doing more harm than good, because consumers will stop inquiring further, even though it is still their duty to inquire and to find out.

D8.5 Title, but not practice?

Why should public policy regulate the title, but not the practice? Merely redressing information asymmetry by identifying regulated providers abrogates the associated responsibility. The message is a mixed one — there is something about this profession that makes it worthwhile for the government to regulate by name, but that something is not so important that the citizen can only obtain it from this one source. Here again Schattschneider’s observation enters — is it merely trivial to regulate title and not activity? If it is trivial, it may also be damaging to government in the longer term.

In the case of the medical practitioner, the answer to the questions ‘what harm can they do? and what redress is there in case of harm?’ is ‘perhaps not much at all, especially if you are dead’. But what about architects? Very few architects actually build houses and other parties are responsible for building houses competently. Architectural plans specify that the ‘builder is to verify all dimensions on site’. A structural engineer certifies strength and stability. Other trades install and certify the services. Architects design, then manage or coordinate the other services. The

'harm' is of a different order. A similar argument can be mounted for other such professions.

D8.6 Summary

When society regulates a profession, to a very considerable extent, it actually constructs the profession. In doing so, society conveys privilege, and, thus, there is good reason to examine the basis for government intervention.

References

- Abbott, A. 1988, *The System of Professions: An Essay on the Division of Expert Labor*, University of Chicago Press, Chicago and London.
- Schattschneider, E.E. 1960, *The Semi-Sovereign People: A Realists View of Democracy in America*, Holt, Rinehart and Winston, New York.
- Stevens, G. 1998, *The Favoured Circle: The Social Foundations of Architectural Distinction*, MIT Press, Cambridge, Massachusetts and London, England.

General discussion

The discussion focused on:

- the reasons for regulating professional services; and
- approaches to regulating professional services.

The reasons for regulating professional services

Legal profession

One participant commented that the legal profession was very heavily regulated at State and Territory level and two important considerations in recent inquiries into the legal profession had supported regulation. One was the consumer interest — that is, the need to ensure that consumers were aware of the services provided, at least at the entry level. The other consideration in the *Access to Justice* report (Law Society of NSW 2000), was the need for efficiency in the judicial system by having qualified and experienced lawyers to appear before the courts.

Medical profession

Another participant drew attention to the plethora of regulation and important changes in the medical profession. The participant considered that deregulation to encourage competition in medical services could have adverse consequences. Consumers, particularly patients in a medical situation, were not always well placed to sort out differences between medical professionals. In the extreme, deregulation could lead to the situation where anyone could become a medical service provider.

An example of the problems of deregulation was medical advertising in New South Wales. Advertising by the impotency clinics had led to very expensive therapies at great cost to the consumer.

Deregulation of professional roles would create more reliance on legal liability after professionals had commenced practice, rather than on a system of checks and balances to ensure the professionals had adequate training before they started to practise. In general practice, for example, the checks and balances included accreditation, the medical services review and ongoing supervisory oversighting by medical boards and colleges. Reliance on legal liability when professional malpractice occurred could come too late.

Financial professions

It was suggested that in financial professions, it seemed that there was not only a branding function, but the professional organisations also obtained substantial revenue from providing certification and education courses for which they were paid. In such cases, it was contended that a conflict of interest might arise between the certification and training roles played by the professional body. It was further suggested that there was a need to ask why such training activities were allowed to compete with universities.

Approaches to regulating professional services

Uniform national regulation

It was commented that the legal profession initially started with self-regulation under independent bodies. In some States and Territories, the legal profession was still fully self-regulating, but this was gradually disappearing as a result of the Competition Policy Review of the profession. Each State and Territory had its own regulatory requirements that resulted in many disadvantages for firms with national legal practices. The biggest challenge for the legal profession was to achieve uniform national regulation, but not to deregulate.

Self-regulation

It was noted that, in a self-regulated model in some professions — such as engineers and accountants — there were many different sub-organisations which vied for membership. There were different levels of qualification and different royal charters, which did not help solve information asymmetry. Self-regulation seemed a cumbersome way of regulating professional services.

Another participant suggested that the medical profession itself was best placed to regulate professional decision-making. There might be a role for the Productivity Commission to examine areas such as pathology, where technology was taking over from professional judgement.

Current regulations that did not allow group negotiations between doctors and area health services or hospitals had created problems for the medical profession. These regulations left it to the bureaucracy to judge the relative worth of individual medical professionals. This was not a decision a bureaucrat could make. While peer review committees could perform such a role, even then it would be fraught with difficulty and danger. It was contended that the previous system, in which

arrangements were industrially negotiated as a whole, was better for the populace and for the employing groups. Judging the worth of individual professionals was difficult, because they all had expertise in particular areas but may not excel in others.

Response

In response, Neil Byron said that the paper sought to examine the role of professional regulation in general in trying to protect the interest of consumers, with particular reference to architects. The paper concluded that the current system for regulating architects was providing information of doubtful quality and largely to those who did not need or use the information. If there was a potentially vulnerable group of clients who needed better information, a self-regulatory system was likely to provide it at least as well as, or probably better than, the existing statutory based systems.

Uniformity of national regulation could be designed and readily achieved, so long as the regulation were non-statutory. Statutory regulation would have to remain State-based.

Heterogeneity of service within a profession created an information problem that professional regulation needed to address. Reservation of title was very low grade information for consumers. A major concern was that the current system of statutory regulation frequently implied to consumers that there was a government guarantee when, in fact, there was not. The statutory reservation of title system for architects appeared to crowd out what could have been a far more informative system. Such a system would result from a self-regulatory scheme, which ensured that those with quality, skills and expertise informed prospective consumers of this.

Reference

Law Society of NSW 2000, *Access to Justice*, Final Report, Sydney, December.

9 Regulating for the safe and effective delivery of pharmacy services

David Greig
ACIL Consulting

9.1 Introduction

This paper reviews the regulation of pharmacies in the light of the recent National Competition Policy (NCP) Review of Pharmacy Legislation, which was completed in February this year.¹ The Review covers a wide range of issues, including advice on drug use, equitable access to pharmacy services, containment of drug costs, the effect of restrictions on competition, alternatives to promote competition and adjustment issues. This paper concentrates on the two most contentious aspects of the regulations:

- controls on location of pharmacies, relocation of pharmacies and pharmacy numbers, which are linked to attempts to control subsidy costs under the Pharmaceutical Benefits Scheme (PBS); and
- ownership of pharmacies, at present largely (though not entirely) restricted to pharmacists.

The NCP Review recommended liberalisation of the first of these regulations, but not of the second.

This paper deals only cursorily with other less contentious issues, such as pharmacy registration and pharmacist qualifications.

The regulations in question have been in place, with variations, for decades. The location regulations are said to be required to contain PBS costs and the ownership regulations are said to be required for quality of professional service. However, the NCP Review (1999, Part A, p. 27) noted that ‘in some cases, their introduction was

¹ The main sources used to preparing this paper were the preliminary and final reports of the Review (NCP Review 1999 and 2000, respectively), and submissions by the Pharmacy Guild of Australia (PGA and PSA 1999) and the Productivity Commission (PC 1999).

a policy response to the community pharmacy industry's lobbying against the entry into the Australian market of an overseas based pharmacy chain'.

The regulations have remained largely intact, despite numerous reviews (eg Ralph 1979, BIE 1985, IAC 1986 and NCA 1996) expressing concern about the anticompetitive effects and effects on drug dispensing costs and drug prices. The main policy change in recent times was an exercise in 1990 to reduce the number of pharmacies, which is said to have reduced PBS costs by over \$200 million over the life of the agreement. Pharmacy interests and government health departments lobbied the NCP Review to retain the main regulations.

As is normal with NCP reviews, the Review was asked to examine pharmacy acts and regulations and determine 'whether these imposed restrictions on competition and, if so, whether any such restrictions are of net public benefit and, if not, whether they should be removed'. The Review was especially asked to look at:

- ownership of pharmacies;
- location of pharmacies to dispense benefits under the Commonwealth PBS; and
- registration of pharmacists.

9.2 Existing regulations

New pharmacy approvals and relocation of pharmacies

The Government has an agreement with pharmacy interests, which restricts the location and number of pharmacies allowed to dispense under the PBS scheme. Without access to dispensing subsidies, pharmacies cannot exist, as 70 per cent of prescriptions are subsidised under the PBS scheme and they account for 60 per cent of pharmacy sales (compared with 40 per cent in 1981). The key restrictions are as follows.

- Pharmacies must generally be at least 2 kilometres apart (changed in May 2000 to 1½ kilometres).
- A new pharmacy must satisfy 'definite community need' criteria, including a catchment area with:
 - more than 3000 permanent residents;
 - inadequate services by other approved pharmacists;
 - prescription volumes of more than 3000 prescriptions per year;
 - general shopping facilities; and

-
- various criteria related to population mix and isolation.
 - Pharmacists may relocate within one kilometre of their existing location every 2 years and may relocate further if the new location is at least 1½ kilometres from another pharmacy and does not leave behind an area of 'definite community need'.
 - Pharmacies may relocate to shopping centres of more than 30 shops with a defined ratio of pharmacy numbers to shop numbers.

Linked to these regulations is a regime under which the Commonwealth Government, through the PBS scheme, defines margins above wholesaler prices for dispensing drugs, plus defined dispensing fees per item. There is also a loading for isolated pharmacies, and extra allowances for medication reviews, asthma and diabetes management and wound care. The fees have been set with reference to average pharmacy costs adjusted to consumer price index (CPI) and other cost movements. Pharmacies are not allowed to compete on price for drugs under the PBS scheme. The Pharmacy Guild sets surcharges for dispensing certain other drugs.

Pharmacy ownership

Pharmacies must generally be owned by pharmacists. There are some grandfathered exceptions:

- a limited number of chains which were in business many decades ago may continue to own pharmacies; and
- a limited number of friendly societies may also own pharmacies.

Other features of the ownership regime are:

- depending on the State, pharmacists may own between 2 and 4 pharmacies;
- defined relatives of pharmacists may also have ownership interests, but not control;
- there are dispensing rights for certain hospitals and doctors in very isolated areas;
- pharmacies may link up for marketing and other purposes (eg Amcal) but no chain ownership is permitted; and
- a pharmacist must be in attendance at all times a pharmacy is open for business.

9.3 Controls on pharmacy numbers, pharmacy location and PBS costs

Review findings

The Review came firmly to the conclusion that the regulations on pharmacy location are anticompetitive. It concluded that ‘it cannot be shown conclusively that the current restrictions are entirely in the public interest’. The Review described the new approval and relocation restrictions as ‘blunt policy instruments’ (NCP Review 1999, p. 73) that protect existing pharmacies from new competitors (some of who may be more efficient or otherwise better). ‘Local communities bear a significant cost of maintaining inefficient pharmacies with little or no opportunity for more efficient competition to enter the local market. Consumers are in effect forced to deal with the established yet unsatisfactory pharmacy ...’. The restrictions lessen the urgency for continuing the process of community pharmacy industry restructuring, ‘they encourage static conditions rather the innovation and positive change’, and ‘... there is no great urge for the proprietors of ... less efficient pharmacies to improve their performance or to review their futures’.

The Review noted that even after the rationalisation of the early 1990s, 72 per cent of pharmacies are still dispensing volumes of under 30 000 prescriptions per year. The Productivity Commission quotes research which concludes that economies of scale only taper out at 80 000 prescriptions per year (PC 1999, p. 36).

The location restrictions have also inhibited normal adjustment to changes in practice — for example, multi-practitioner medical centres, which would be a natural location point for a pharmacy or the increasing proportion of shopping done in large shopping centres.

The four earlier reviews mentioned in section 9.1 all concluded that the regulations had the effect of increasing PBS costs. PBS costs are of concern in overall health policy because they are growing faster than health expenditure in general, and risk crowding out other forms of health expenditure.

Control of PBS costs

The Review, along with the submission by the Productivity Commission, suggests that the government could better control PBS costs by various forms of competitive pressure. For example, pharmacies could tender for PBS business (although it was concluded this would be clumsy) or the government could introduce a scheme now working in British Columbia, whereby each pharmacy is free to set its dispensing

fee subject to an upper limit, and therefore the price of prescriptions. A patient copayment provides an incentive to shop around.

Advertising of prices can back up a regime of this sort. (Advertising is permitted now, but there is little point when PBS prices are controlled and when there are pricing setting arrangements for non-PBS pharmaceuticals.) Some consumers would switch to the more efficient pharmacies offering prescribed medicines for a lower price. Normal competitive forces, the same as those applying in most areas of the Australian economy, would then act to rationalise the industry — efficient pharmacies would expand, inefficient pharmacies would close or would become more efficient. The quality of service would still matter, as it would be a competitive factor.

Pharmacies in rural and regional areas

Perceived wisdom was that the restrictions on location and numbers were necessary to protect rural pharmacies. Experience has shown this to be incorrect. The Review concluded that there was little such benefit. Even the Commonwealth's own submission to the Review noted that 'the existing controls ... have only been effective at the margin in maintaining access in rural and remote areas'.

Pharmacy services are a derived demand — there will not be demand for them where there is not the population. Population changes happening in rural and regional Australia for reasons having nothing to do with pharmacy, such as declining commodity prices and changing farm technology, mean that pharmacies in some isolated areas have become or are becoming uneconomic.

The best policy is to face this reality and deal with the consequences positively, for example, by:

- encouraging mail order pharmacy — mail order firms could compete in part on their willingness to provide telephone health counselling services;
- allowing more doctors in isolated areas to dispense medicine — there is a general presumption against allowing this because doctors would have an incentive to over-prescribe, but this can be tempered by monitoring prescribing patterns; and
- offering subsidies to rural pharmacies — there is a rural loading in the PBS subsidies, and the Government recently announced additional assistance, similar to that announced for doctors.

9.4 Pharmacy ownership

The case for pharmacist ownership

Besides the reasonable requirement that a trained pharmacist must be on duty at all times that a pharmacy is open, there is a longstanding requirement that, with a few exceptions, pharmacies must be owned by pharmacists. There are also restrictions on the number of pharmacies which an individual pharmacist may own (between two and four depending on the State), on the ownership structures of the pharmacy businesses, and on related financial structures (so that a non-pharmacist may not control the business).

The NCP Review has set out the objectives behind the restrictions (NCP Review 2000, Part A, pp. 27–36). These objectives, upon closer examination, all fall short. The objectives are summarised below in italics together with criticisms.

- *Promoting the integrity of pharmacy as a professional activity as opposed to a commercial activity.* These two need not be in conflict — a well-run pharmacy will attract more customers than a badly run one. The integrity of the pharmacy is protected by the requirement to have a trained pharmacist on duty at all times that the pharmacy is open.
- *Industry-wide awareness of professional pharmacy objectives as well as commercial objectives.* This has nothing to do with ownership *per se*. Trained pharmacists on duty in pharmacies can be expected to pursue professional objectives, and the Pharmacy Guild or other associations can pursue industry wide objectives. Non-pharmacist owners could support such efforts — as do the friendly societies.
- *Keeping pharmacy businesses small enough to facilitate the close personal supervision of the professional operation by the proprietor pharmacists.* It is not clear what is achieved by this if a trained pharmacist must, in any case, be on duty at all times the pharmacy is open. In any case, pharmacist proprietors are not always able to provide close personal supervision — they are allowed to own up to four pharmacies, some of them are retired, and some pharmacies even now may be owned by non-pharmacists with no evidence of ill effects.
- *Making the proprietor of a pharmacy business professionally and directly accountable to regulatory authorities.* The duty pharmacist must, in any case, be accountable for his or her professional activity. It is not clear what else this additional layer of accountability that is not found in other health areas (eg doctors and dentists) achieves.

-
- *Underpinning the ease of Australian's access to community pharmacies wherever they live.* As discussed in section 9.3 above, this is not being achieved. It could be argued that the combination of ownership and location restrictions is slowing the rate that pharmacies in the smaller rural towns, and smaller suburban shopping centres, are losing ground to pharmacies in larger rural towns and larger shopping centres. However, as argued elsewhere in this paper, other regulations are having the opposite effect. In any case, ownership restrictions are a blunt instrument here — mail order pharmacy, dispensing by rural doctors and explicit subsidies would be more targeted policies.
 - *Assisting the efficient allocation of scarce public resources on pharmacy services.* This is not achieved — the opposite is the case. Regulation of pharmacy location and the cost-plus PBS regime protect high-cost pharmacies from competitive pressures. Regulation of pharmacy ownership limits the flow of retail business skills to pharmacies.
 - *Improving the capability to link community pharmacy ... to overall health care provision and multi-disciplinary service provision.* Ownership restriction is not necessary for this, as the duty pharmacist has been trained for such activities. And the location restrictions can inhibit such links, by sometimes preventing establishment of pharmacies in multi-practitioner medical centres.
 - *Protecting independent pharmacy businesses from perceived 'unfair competition' and the market dominance from large pharmacy owning corporations and chains and, in some jurisdictions, friendly societies.* This may be the real reason for the longstanding pressure to maintain the ownership restrictions. It is an argument that is clearly incompatible with NCP and protects inefficiency.

A later part of the Review picks up themes of submissions from pharmacy interests and elaborates on some of the above statements as follows (again the Review's wording is in italics followed by this paper's criticisms).

- *It is important that, in the interests of public safety and the best use of medicines, pharmacies focus on their professional responsibility in ways that recognise the effects of information asymmetries between the pharmacist and their consumers.* The duty pharmacist is trained to do this. A pharmacist ownership layer on top is redundant, unless the duty pharmacist is behaving unprofessionally, for which there are other remedies. Other professions, such as doctors and dentists whose activities also affect health and public safety and involve information asymmetry, do not have this extra layer.
- *... non-pharmacist proprietors would not be able to be made readily accountable (to the regulatory authorities) without a major and potentially costly readjustment of the regulatory infrastructure.* This is incorrect. Already

the non-pharmacist owned pharmacies owned by friendly societies and grandfathered chains are satisfactorily accountable to the regulatory authorities.

- *There is a net public policy benefit in promoting pharmacist proprietorship and control in the community pharmacy industry to maximise the distribution of pharmacies into the wider Australian community, to encourage proprietors to have a more direct relationship with a local community, to promote ... the wise use of medicines, and to ensure the maximum possible social and geographic reach of the community pharmacy network.* Pharmacist ownership will not ensure the survival of pharmacies in areas where demand is insufficient to support them. This argument appears to be code for inhibiting expansion of the more efficient pharmacies. However, if the more efficient pharmacies attracted business away from marginal pharmacist-owned pharmacies, that would be the result of individuals' choice, because they were attracted by lower prices or better service. Proprietors who do not work in their pharmacies are unlikely to have a more direct relationship with a local community — the duty pharmacist usually achieves this relationship.
- *The PBS ... is predicated on the stability of this distribution network and while this may mean new entrants are discouraged, undue changes to that network could mean the costlier and less effective delivery of PBS medicines.* This is not plausible. Any sound administrative system can cope with changes in suppliers. As argued elsewhere, current policies increase the cost of delivery of PBS services because of the suppression of competitive forces.
- *There is a significant net quantitative benefit to the community from reduced and unnecessary expenditure due to the sensible intervention of professionally oriented pharmacies in the health care cycle, such as in reducing pressure on general practitioners in the treatment of minor ailments. Having pharmacists owning as well as managing pharmacies helps to reinforce this professional role and culture.* The trained pharmacist who must be on duty in any pharmacy is able to carry out this undoubtedly valuable service. Having an extra layer of pharmacist ownership on top is redundant.

There was an additional argument in earlier Pharmacy Guild papers, not picked up in the Review, which suggests that commercial pressure in US pharmacies for pharmacists to dispense an unreasonable number of prescriptions per hour has led to a noticeably higher level of incorrect prescription (with obvious health risks) than in Australia. The nature of the evidence provided was anecdotal. The Productivity Commission's comparison of performance in Australia and in United Kingdom, the United States and parts of Canada suggests there is no systematic deficiency in the latter three countries, where pharmacist ownership is not required. Even if there were such a problem, pharmacist ownership does not guarantee against it: individual pharmacist owners could also be 'greedy'. The issue is best addressed by

direct measures — for example, a maximum prescriptions/pharmacist ratio or performance monitoring — rather than the blunt instrument of ownership restriction. There is no constraint to imposing statutory codes of practice on pharmacy owners. Retailers, for example, are already bound by such codes in areas such as food handling and public health (NCP Review 2000, Part A, p. 44).

A related issue discussed in a 1997 Pharmacy Guild paper was that of the tort system, and the minimal use made of it in practice to address negligence claims. This would suggest a case for compulsory professional indemnity insurance rather than ownership restrictions.

The case against the ownership restrictions

The costs imposed by such restrictions are well known and underlie the State and Commonwealth Government agreement to implement NCP:

- *Barriers to greater efficiency.* The powerful combination of:
 - ownership restrictions;
 - location restrictions;
 - the average-cost basis of PBS subsidies; and
 - the ban on price competition for subsidised medicines and constraints on price competition for non-prescription pharmaceuticals

have the effect of protecting inefficient or poor-service pharmacies from the normal disciplines of competition, contrary to the intention of NCP. If service is poor, the customer must either put up with it or put up with the inconvenience of going to a pharmacy further away. Prices of both PBS and other pharmaceuticals are higher than they need be because of the absence of competitive discipline on costs. The opportunity to combine a pharmacy with other retail businesses — for example, a pharmacy inside a supermarket or a department store — is lost and with it, the opportunity to share costs, overheads and retail skills.

- *Protection of poor service.* Submissions from various consumer groups to the NCP Review provided evidence of mixed quality of service. The Review said ‘consumer evidence to the Review tends to suggest that many pharmacies’ emphasis on quality of service is relatively indifferent and patchy’, and ‘a poorly performing pharmacist proprietor can be comfortable in the knowledge that he or she is not exposed to competition of more aggressive and commercially able non-pharmacist competitors. They are also reassured that the PBS and its pharmacy location rules help to buffer them against new pharmacists entering their immediate catchment area’.

-
- *Restricting sources of industry innovation.* The Review notes that pharmacist proprietors tend to be older, predominantly male and many are no longer in active practice. The restrictions make it more difficult than necessary for younger pharmacists to own pharmacies and also restrict the access to business and retail skills, many of which are found largely outside the pharmacy sector. The Pharmacy Guild has programs that attempt to compensate for this, but they are not a substitute for the real thing.
 - *Constraints on consumer choice.* Besides denying consumers access to lower cost pharmaceuticals, ownership restrictions also deny them the choice between different levels and types of service. For example, as in other areas of retail, some customers would prefer large stores with low prices and a wide choice, while others would prefer smaller stores with higher prices and personal service. Some would prefer the convenience of pharmacies located in medical centres.

Pecuniary interests

A decision to restrict pharmacy ownership forces consideration of pecuniary interests to ensure that other commercial arrangements do not undermine the control of the pharmacist over the activities of the pharmacy.

The Review gets into a muddle here, convincing itself that minority shareholdings pose a risk (they do not if small or diffuse), but accepting continuation of traditional pharmacy debt financing often provided by pharmaceutical companies and of preferred supplier status with pharmaceutical companies. Both of these can significantly affect the way the pharmacy business is managed. Short of defaulting, the obligation to service debt is a more immediate and pressing one than servicing equity, unless the debt provider is flexible — thus, placing an obligation to look after that debt provider's interest. A pharmacist caught up in debt may have more difficulty escaping if the only potential buyers of his/her pharmacy are other pharmacists. Preferred supplier status may not be fully compatible with a pure interest in the health and well-being of customers. At another stage, the Review talks of joint ventures and naively suggests that 'if the proprietor fears a loss of control of their arrangements to the "host" company, then they will presumably not enter into a joint venture arrangement ...' (NCP Review 2000, Part A, p. 64).

These contortions would not be necessary if there were no restriction on pharmacy ownership.

Contradictions within the Review

There are a number of contradictions within the Review itself, which undermine its own conclusion.

- Pharmacists are praised as ‘... professionals whose high training enables them to trust their personal judgement and make all manner of decisions carefully and analytically’. It is not then clear why these people need another ownership layer on top to ensure they behave correctly.
- The Review admits that non-pharmacist owners can behave correctly. ‘This is not to say that non-pharmacist proprietors are not capable of building such professional culture into pharmacies they operate. Friendly Society and surviving corporately-owned pharmacies in Australia do promote professional values in their operations’. The Review argues that a pharmacist proprietor is more likely to favour the customers’ health needs over business interests. But, this presupposes that other owners are not concerned about their reputations (which in a retail industry reliant on repeat business they normally will be) and that pharmacist owners are not under commercial pressure (as discussed below, some are).
- Friendly Society pharmacies are already accountable to the Pharmacy Boards through their own (non-pharmacist) boards of management. The Review says ‘in its consultations with regulatory authorities, the Review found general satisfaction with the working of these arrangements. There is no reason to believe that the same principle could not be extended to non-pharmacist individuals and corporations’. The Review also states that the exceptions to pharmacist owned pharmacies ‘appear to work well as competently managed and professionally sensitive pharmacy businesses’.

The Review claims that large financially strong owners would be better placed than pharmacist owners to contest regulatory outcomes with which they were unhappy through protracted litigation. However, the non-pharmacist owners would become a significant force only if there was deregulation of entry — hence no ‘regulatory outcome’ to be unhappy with.

Perhaps one of the most telling arguments against requiring pharmacist ownership is found in the Review itself:

That non-pharmacist proprietors are capable of providing safe and competent pharmacy services suggests that allowing new non-pharmacist proprietors would not necessarily destroy the local pharmacy network and infrastructure to which Australians have become accustomed. It certainly would not be in the commercial interests of non-pharmacist proprietors to expose themselves to the risk of loss of income or profit, or litigation, due to their pharmacies being unsafe or incompetently run. (NCP Review 2000, Part A, p. 46)

Given the contradictions, it is not surprising that the Review had misgivings about its conclusion in favour of the status quo. It refers to the conclusion having been reached ‘on balance’, substantial parts of the report dwell on the arguments for dropping the restriction and concern is expressed about the complaints from several consumer organisations, with a plea for pharmacist proprietors to heed their messages (NCP Review 2000, Part A, p. 59).

9.5 Evaluation

The Review presents a generally sound evaluation of the regulation of pharmacy location and the related cost-plus approach to PBS subsidies and PBS price fixing. Despite a preponderance of submissions from pharmacy interests and governments favouring the status quo, the Review applied the principles of NCP (and the same thing — economic common sense), concluding that the regulations were not achieving their objectives and that competitive processes would work better.

The Review’s consideration of ownership restrictions was not of such quality. Although it devoted some space to the arguments for removing the restrictions, it concluded ‘on balance’ in favour of the status quo. Throughout the pages of the Review that cover the reasons for this, and throughout the Pharmacy Guild submission, one is left searching for a solid reason — as if the word processor had somehow missed some key paragraphs. Neither document really explains what is achieved by superimposing a layer of pharmacist ownership on the professional activities of trained duty pharmacists (whose competence is fulsomely praised). This section fails as a NCP review: it establishes that the regulations ‘impose restrictions on competition’, but does not show that ‘such restrictions are of net public benefit’ and, hence, does not make a case for keeping them. One is left wondering whether the Review had been ‘got at’ to produce the conclusion it did.

Nevertheless, the Review’s recommendations, if implemented, would produce some positive results.

- Pharmacy location would quickly rationalise to line up with demand — more pharmacies in large shopping centres and in medical centres, closure of the more marginal suburban pharmacies, and a possible saving of some otherwise doomed rural pharmacies thanks to cost reductions from co-location with other business(es).
- Despite the ban on minority shareholdings, supermarket and chain pharmacies would establish through the provision to pharmacists of retail space, debt financing and assistance with retailing skills — all permitted under the Review’s recommendations.

-
- Certain pharmacist owners whose businesses are worth little because of poor location would get windfall gains, as they would more readily be able to move. Other pharmacists whose businesses' goodwill value is the result of protection rather than service competence would make capital losses when the protection was removed. A few of the more astute pharmacist owners would make large windfall gains by becoming the 'figurehead owners' for large numbers of pharmacies (the limits having been removed) in supermarket or other chains.
 - There would be increasing competitive pressure on PBS and other margins.

Overall, there would be an improvement in efficiency, a better harnessing of retail and financial skills, lower PBS costs and prices, lower non-PBS pharmaceutical prices and greater consumer choice of types of pharmacy. These gains would outweigh the peculiar effects on the fortunes (economic rents) of individual pharmacists. However the adjustment would work more effectively and completely, without the windfall gains, if the ownership restrictions were also removed.

9.6 Government policy

The reaction of the Commonwealth Government so far has not been encouraging. The Review was completed in February 2000 and in May the Government announced it had reached a new agreement with the Pharmacy Guild. The announcement referred to substantially higher payments to pharmacists and the retention of location rules, apart from a reduction in the minimum distance between pharmacies from 2 to 1½ kilometres. High-convenience location of pharmacies in multi-practitioner medical centres continues to be forbidden if it breaks the 1½ kilometre rule. A Government spokesperson is reported as saying that the Government did not always accept NCP reviews.

The Government is continuing the cost-plus PBS mechanism, and is continuing to try to temper its inevitable consequence — that of rising costs — by direct controls on pharmacy numbers and location. This does not solve the cost-plus problem, because costs are based on those of an average pharmacy, not of an efficient pharmacy, and the regulations themselves reduce the pressures for pharmacies to become more efficient. The Commonwealth has trapped itself into a circular logic, where one flawed policy causes another that in turn undermines the first.

Despite the liberalisation of the Australian economy over the last few decades, which has underpinned its solid economic performance, a few sectors — taxis, newsagents, certain professions and pharmacies — continue to be regulated in a way reminiscent of Australia in the depression war years of the 1930s and 1940s, India in the centrally planned 1960s and pre-1989 Eastern Europe. The vested

interests lobby hard to keep their privileges, while the costs of inefficiency and relative lack of innovation are borne by the rest of the (liberalised) economy.

Such an outcome is to be expected, given the mechanism by which the Government makes its policies. Instead of allowing competitive forces, from the better pharmacies and discerning customers, to shape the sector, it ignores the consumers and does deals with the suppliers. ‘Regulatory capture’ is evident. The Government has created a cartel and enforces it. The result is that taxpayers and customers pay more for the dispensing of pharmaceuticals than they should.

References

- BIE (Bureau of Industry Economics) 1985, *Retail Pharmacy in Australia — An Economic Appraisal*, Research Report no. 17, AGPS, Canberra.
- IAC (Industries Assistance Commission) 1986, *Pharmaceutical Products*, Report no. 382, AGPS, Canberra, April.
- NCA (National Commission of Audit) 1996, *Report to the Commonwealth Government*, AGPS, Canberra.
- NCP (National Competition Policy) Review 1999, *National Competition Policy Review of Pharmacy*, Preliminary Report, (Mr Warwick J. Wilkinson, Chair), Council of Australian Governments, 4 November.
- 2000, *National Competition Policy Review of Pharmacy*, Final Report, (Mr Warwick Wilkinson, Chair), Council of Australian Governments, 18 February.
- PC (Productivity Commission) 1999, *Productivity Commission Submission to the National Review of Pharmacy*, Canberra, November.
- PGA and PSA (Pharmacy Guild of Australia and Pharmaceutical Society of Australia) 1999, *Submission to the National Competition Policy Review of Legislation*, Barton, Canberra, July.
- Ralph J.T. (Chairman) 1979, *Pharmaceutical Manufacturing Industry Inquiry*, Report, AGPS, Canberra, August.

Discussant — *Helen Lapsley*

University of New South Wales

There is much to agree with in David Greig's paper. The paper provides a useful discussion of the existing regulations and surprising aspects of the new Pharmacy Agreement. These comments are, however, a bit more controversial.

D9.1 Defining the relevant issues

The paper correctly separates the issues of pharmacy control, numbers and locations, and pharmacist ownership. These issues have been rolled into the same issue in policy considerations.

It is important to note that pharmacists themselves are already regulated through professional registration boards. The Pharmacy Agreement relates primarily to the commercial and retail aspects of pharmacy business and the arguments for commercial privilege and subsidy. However, many professional aspects of pharmacists have also been rolled into the Agreement. The professional registration issues are quite obfuscated with the commercial aspects of pharmacy business. Given the contradictions and qualifications in the NCP Review, it is difficult to objectively support this level of ongoing regulation. Perhaps it is an example of regulatory failure, and it is probably useful to consider that public interest and public protection may not coincide.

D9.2 Control on pharmacy and PBS costs

In relation to the NCP Review, it would be useful to establish (if possible) a relationship between pharmacy numbers and total PBS costs. The Review is remarkably short on hard data. This is surprising given that the Government is the primary payer, is involved in negotiating the dispensing fees, and is also aware of those pharmacies that have been grandfathered or are owned by friendly societies. It would be very useful to evaluate rigorously whether those pharmacies behave any differently from the others that the profession and the Pharmacy Guild insist be owned by pharmacists.

D9.3 Ownership restrictions

The ownership of pharmacies by pharmacists is in direct contrast to the way in which other professions are organised. For example, doctors do not need to own their own medical practices, and, at least in New South Wales, are positively

discouraged from hospital ownership. Thus, it is quite interesting that pharmacist ownership is officially supported.

D9.4 Dispensing fee and drug costs

It would be useful if David Greig's paper could separate the arguments relating to the dispensing fee charged by pharmacists from the costs of prescribed drugs and their mark-up. Both the NCP Review and David Greig's paper seem to suggest that these functions are the same. However, the reality is that the major interest of government in relation to pharmacy practice is the size of dispensing fees because there are, of course, other mechanisms to access prescribed drugs.

Pharmacists do not initiate the supply of prescription drugs. It is the amount of dispensing fees that should be addressed, not the actual numbers of prescriptions that pharmacists dispense, because that is not actually within their control.

D9.5 Prescription drugs

Another useful aspect to consider is the numbers of pharmacy-only drugs and how those are handled within pharmacies. One of the arguments has always been that pharmacies are special and, hence, the non-prescription pharmacy-only drugs should be provided only by pharmacists. However, anecdotal evidence that consumer groups have presented to the NCP Review suggested that this argument does not hold up.

D9.6 Rural and remote pharmacies

It would be of interest to establish whether there are any data relating to the current accountability mechanisms, to confirm the differences (if any) between the existing and non-pharmacist owned pharmacies.

It would be useful to look at the proposed provider number analogy for general practitioner location, and compare that with the subsidy for rural and isolated pharmacy locations. That is one area where there are genuine public interest arguments, and the subsidy paid to those pharmacists reflects that. But such legitimate concerns do not justify the protected regulatory environment for the *entire* industry, given that access problems apply to only a small percentage of the retail pharmacy sector.

The paper questions whether, in rural and remote areas, it is appropriate to continue to separate prescribing and dispensing roles, as is law in Australia. However, in

reality, there is an even greater shortage of rural General Practitioners than rural pharmacists. It is difficult to imagine any support for an additional role for rural and remote doctors to prescribe and dispense. This is a very interesting and separate international issue, which many countries have struggled with.

There is a need to look at the issues of genuine access for rural and remote areas and to address the contradictions on this within the NCP Review.

D9.7 Summing up

The new Pharmacy Agreement has become more complicated over time, involving ever more concessions. The results attest to the effective lobbying that raises definite questions that still remain unanswered about the public interest. It is probably time to take a more rigorous look at this whole mishmash, and David Greig is correct, the NCP Review did not really do that.

General discussion

The discussion focused on:

- the NCP Review;
- the regulation of pharmacies; and
- rural and remote pharmacies.

The NCP Review

Aspects of the NCP Review were discussed. One participant suggested that the Review could send the wrong signals to other professions. The Review did not establish the need for pharmacist ownership of pharmacies, since requiring professional pharmacists to be on hand in the pharmacy would ensure that the required functions would be met. Evidence on adverse consequences of non-pharmacist ownership was lacking in the NCP Review. For example, in the United States, non-pharmacist drug stores and chains appeared to be fairly competent in the roles of providing advice and dispensing drugs. The participant thought that there was a need to educate the community that the public interests may be better served through a different model. The lost opportunity for reform reflected strong sectional interests.

Another participant commented that although it had been referred to as a NCP review, the Review did not actually satisfy the National Competition Council (NCC) requirements of the review process. Thus it was unclear how binding the findings of the NCP Review would be. Deregulation may yet occur. Changes in technology and community behaviour may require pharmacies to adjust to the new environment, and existing pharmacy regulations may actually hinder pharmacies in adjusting to these changes.

It was asked whether trade issues had put pressures on the regulatory arrangements of pharmacies. New technology had made it possible for consumers to gain access to pharmacy services overseas, especially for types of drugs that were not urgently needed. Foreign firms were also seeking to provide a wide range of products, including pharmacy products, in Australia and this could put pressures on the way pharmacy regulation was managed. In response, David Greig noted that he did not find any direct trade policy issue at the inter-government level. However, development of mail-order pharmacy was certainly an issue in regulatory arrangements. Over time, these issues would have some impact.

The regulation of pharmacies

A participant agreed with the discussant that it was not the pharmacists that were driving drug costs, but rather the prescribing patterns of doctors. In light of this, prescribing patterns should be looked at and dispensing fees of pharmacists should be subject to competition.

However, another participant pointed out that it was a misconception that pharmacies were prevented from discounting their dispensing fees. The participant stated that pharmacies could completely discount their dispensing fee and mark-ups associated with the general patient contribution.

Another participant stated that the Pharmacy Agreement did not contribute to increased budget remuneration for pharmacists. There were considerable savings to the forward estimates of the Commonwealth Government coming out of the review of that agreement. The available savings were being redirected, at the margin, towards rural and remote pharmacies. Although this might not be the best way to address the issue of rural pharmacies, it was not an increase in money going to pharmacies.

In response, David Greig agreed that the forward estimates had shown some savings. However, the Pharmacy Guild's summary of part of the Agreement indicated that the subsidy savings were going back to pharmacies in several ways. An extra lump sum payment of \$11 million was to be provided to pharmacies. The Agreement also provided for fee increases, quicker payment of the subsidies than before, financial assistance to rural pharmacies, and a funding of \$10 million to encourage export of information technology in pharmacy.

Rural and remote pharmacies

One participant drew attention to the fact that the ownership regulations may be having adverse impacts on rural and remote areas. In rural areas that were facing difficulties, one way for businesses in the area to become viable was to co-locate, combining into a critical mass business. For pharmacies, it was claimed, such business practices were prevented by the ownership restrictions. Consequently, pharmacies could become less viable and the critical pharmacy service required for the survival of a rural area could be lost. Once this process happened, the whole infrastructure of the area could break down. Advocating pharmacy regulation to protect rural towns could have the opposite effect.

In response, it was pointed out that, while pharmacies could not be owned by other business, they could certainly co-locate. However, it was argued that the critical mass of the business, and not co-location, was the critical issue.

PART E

REGULATING TO
ATTAIN SOCIAL GOALS

10 Is market-based regulation the best way to regulate community services?

This chapter summarises a debate between Alan Moran, of the Institute of Public Affairs, and Robert Fitzgerald, of the NSW Community Services Commission, about whether market-based regulation is the best way to regulate community services. Each participant presented their case, before responding to the points put forwarded by the other. A general discussion followed.

10A The case for market-based regulation

Alan Moran

Institute of Public Affairs

10A.1 Different types of regulation

Liability law and regulation

Liability law and regulation are often thought of as alternatives.

Liability law is based on a notion of fairness in commercial dealings that is adapted and refined through the decisions of judges in contested cases.

Regulation often starts with the system of law based on fair exchange and, in a sense, codifies it. Often, regulation simply gives a statutory expression to what has been common practice. But, sometimes it tries to re-fashion the law by seeking out greater equity and it is here that distortions are created. Regulation also has less flexibility than liability law, which is inherently organic and adaptable to changing circumstances.

Liability law operates by penalising those who do not offer the value promised. Regulation often tries to prevent the unwanted from happening in the first place.

In many respects, the choice between liability law and regulation reflects a choice of how the economy should operate.

The private enterprise system is based on people deciding for themselves how to spend their money, and on producers competing for the consumer's dollar. In order for producers to be successful, competition forces them to bear down on costs and ensure their offerings are attractive. The private enterprise system is based on atomised decision-making. This also allows the users to decide price/quality tradeoffs without others imposing their undoubtedly well-considered tradeoffs on the rest of the community.

Intuitively, the regulatory approach has some merit:

- its preventative, rather than reactive, approach appears to give bloodless benefits;
- the experts, by definition, appear to be better placed than the average decision maker; and
- there are economies in search activity — regulations bring minimum standards, allowing consumers to avoid time-consuming reviews of different providers.

But, in practice, there are many deficiencies of the regulatory approach:

- standards might not match the consumers' wishes and uniformity might not be appropriate;
- regulated supply of goods and services is susceptible to capture by producers, thereby introducing excessive costs;
- the standards often bring weak incentives to adapt to shifting demands; and
- the need to have goods or services approved prior to being allowed to trade, which is frequently the case in regulated systems, brings paper-burden costs and often either abuse or inflexibilities.

Liability law performs better, but is not always the best approach in all circumstances. Indeed, under the custodianship of liberal jurists in the United States and elsewhere, the law wandered from the traditional liability that sought to pursue efficiency by allocating responsibility to the party able to achieve it most cheaply. Liability has been shifted onto the supplier. There have been many peculiar lawsuits, like the burglar who fell through a plate glass roof and successfully sued, or the motorist who inflated his tyres to several times the pressure recommended on the rim and was able to collect after the inevitable mishap.

But this change was only in part because lawyers started to impose their brand of justice, rather than interpret the law. It was also because modern life meant the old

principles of *caveat emptor* could not be applied when the buyer could not easily examine goods prior to purchase.

Output versus input regulation

In principle, liability law represents a complete coverage of market mechanisms. In comparison, output regulation is a partial application of these mechanisms, as it specifies the outcomes and leaves the operators to seek out the lowest cost means of supplying these.

Like the dichotomy between liability law and regulation, the choice between output and input regulation is superficially straightforward, but may be less so on reflection.

Many economists learnt from the more notorious examples of input regulation — laws which specified the size of fishing boats, their engine capacity, their net size and the times they might fish. The outcome was massive overcapitalisation designed to maximise the catch and little concern about resource sustainability. Building regulations used to specify the precise details of a door support and made it extremely difficult for anyone that developed a superior system to comply with the regulations. Economists who suggested alternatives to these command-and-control regulations were pilloried as dreamers ideologically fastened to a free market, but few people nowadays would deny the superiority, and inherently greater flexibility, of output controls.

That said, such controls are not always easy to devise.

10A.2 The community services to be addressed

While the concept of community services includes hospitals and police stations, this paper specifically addresses childcare and aged-care services, as these services explore the depths of the arena for market solutions, because of the parties with which they operate. Children are not capable of exercising informed choice, while the infirm may have impaired faculties.

The lack of full information and capabilities on the part of the consumer does not rule out having a liability-based, rather than a regulatory-based, system. In reference to car repairs, share purchase decisions, computers or a great many other goods and services, markets work because the consumer either treats the supplier as their agent (and a poor reputation will starve the agent of business) or because an external

agent is used. This is because the onus for poor outcomes has moved from buyer to seller, under the pressure of increased complexity of modern life.

In any event, these markets generally work efficiently with little regulatory control.

Similarly, the use of output-based measurements is now viewed as superior in areas, such as education and hospitals (with casemix — that is specifying fixed payments for well-defined treatments like hip replacements). Casemix, and the success of US Health Maintenance Organizations (HMOs), demonstrate that there is every reason to expect health services to work better under an output-based system, rather than one regulated through inputs. The HMOs demonstrate the benefits of moving still further along the true market path.

10A.3 Childcare

Developments in provision and costs

Childcare centres show the deficiencies of controls, especially input controls.

Over the past twelve years, there have been three very significant directional changes in the provision of childcare.

- First, there has been a shift to have the Government pay a greater share of costs. Commonwealth spending in dollar terms rose fourfold during the 1990s and the number of children being cared for appears to have almost doubled. The Commonwealth presently furnishes about 60 per cent of the revenue of childcare centres.
- Second, there has been a remarkable shift from public (community) to private provision. There are now about three times as many private as government or other non-profit childcare facilities.
- Third, there has been an avalanche of new regulations.

As a result of increased regulation, real costs increased by 37 per cent in the case of community centres, 34 per cent for private centres and 12 per cent in the case of family daycare from 1991 to 1998.¹ These costs are still increasing, as childcare workers are trained and some existing workers with no formal qualifications are ‘grandfathered’.

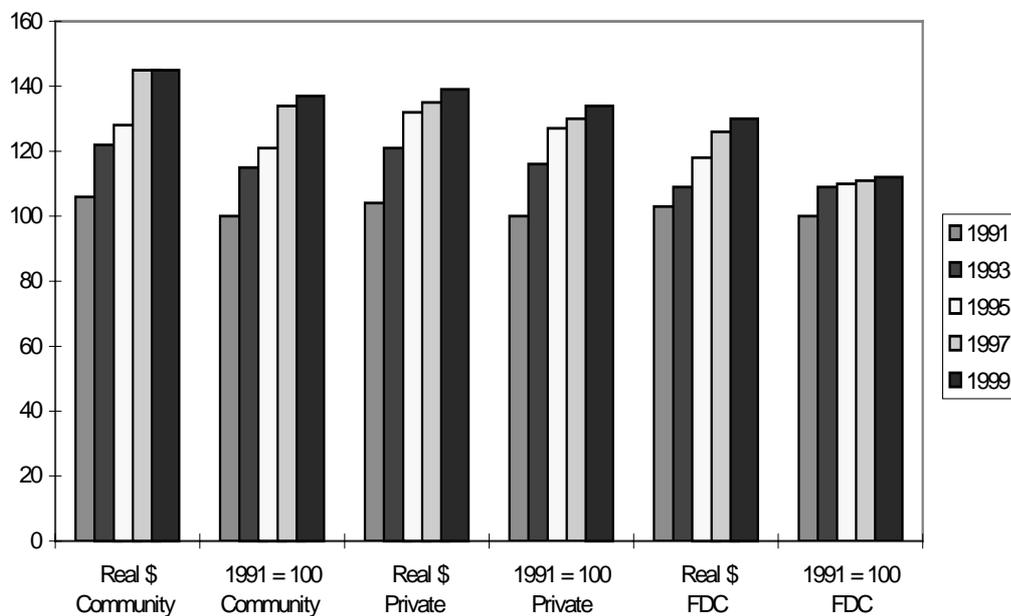
¹ Family daycare is in home-based facilities with only three to four children being cared for.

The relatively low increase in family daycare costs occurred despite the imposition of an administration levy for the first time in 1997. At the start of the decade, family daycare cost almost the same as other systems, but by 1998 daycare was 26 per cent cheaper than community centres and 21 per cent cheaper than private centres. The relative increase in community centre fees is partly a result of the discontinuance of centre specific subsidies and their replacement by subsidies to parents. These trends are illustrated in figure 10.1.

Commonwealth funding has increased considerably over recent years and was over \$1 billion in 1997-98. About 42 per cent of children are fully financed by government assistance (which provides some 70 per cent of the costs) and about 33 per cent receive partial support. Hence, government funding amounts to just over 60 per cent of the total revenues of childcare centres.

Government assistance has increased at twice the rate of the consumer price index, but has failed to keep pace with the fee increases. There is a growing gap, in money terms, between the maximum fee payable and the average price for childcare services. But, the service is clearly competitive. The observation that fees have increased significantly does not suggest market failure.

Figure 10.1 Fees charged by childcare centres



FDC: family daycare

Data source: FACS (1999).

Private centres had assumed the most important role in terms of provision by 1998 and have almost trebled in places. Community-based centres had increased by only 31 per cent and family daycare by 50 per cent.

The authorities may wish to nurture the family daycare sector, given the relative performance (recall that their cost increases were only one third of other centres). After all, the service was demonstrating consumer appeal by winning market share.

However, the increased market share of family daycare centres can be attributed to lower operating costs resulting from less onerous obligations to have child-specific facilities and staff training requirements. Hence, regulatory authorities are finding that this class of service is more attractive to the market than they think is good for the consumer. As a result, the low cost of family daycare is largely a thing of the past (though the statistics do not yet reflect this). As of mid-1999, New South Wales, Queensland, and Western Australia have regulated the family daycare sector, and other States have also implemented some form of regulation (such as minimum standards and staff-child ratios).

A closer look at costs

For the purposes of this discussion, the most important feature of childcare is the cost increase. The increase in fees can hardly be attributed to a natural increase in costs. The basic costs involve:

- labour skills, which, since time immemorial, have come naturally to people; and
- facilities that most families have if they are bringing up children.

Impact of credentialism

The increase in costs created by new regulations must be attributed largely to the credentialism that is specified. Rearing and caring for the young is something that requires no qualifications. It is a natural function of all creatures. Millions of generations have demonstrated that it requires no scholastic preparation.

But there has been a progressive extension in the requirements to have qualified staff employed in the centres. Commonly, when children are present, state regulations require a minimum of two such staff in attendance at all times.

Some may argue that childcare is a form of education and that qualified staff are equally important in this service as in schools themselves. If this is the case, the preferred approach is to act directly and lower the school attendance age.

The adverse impacts of credentialism can be crystallised under three headings.

- First, there is the needless cost of acquiring the qualifications. If two years are taken to obtain the relevant diploma, and the normal working period is twenty years, society incurs a front-loaded cost increase of 10 per cent. The hapless consumer or taxpayer pays for that cost.
- Second, and related to this, the regulations will directly increase wage costs. According to the Victorian Regulation Impact Statement (p. 31), credentialled workers are paid 29-50 per cent more than other workers. This will be reflected in higher fees.
- Third, it denies the opportunity for some of the least-privileged members of society to use their skills. The qualifications will prove too onerous for many willing carers. For others, the need to obtain the credentials and forgo paid employment will prove to be financially forbidding. The impact is likely to be on individuals, especially women, from poor backgrounds, including indigenous Australians, and migrants.

Staff to children ratios

There have been progressive increases in the required ratio of staff to children in all childcare centres. A better approach is to abandon mandatory staffing levels and replace them with a requirement that all centres prominently display up-to-date information on the number of children, the number (and qualifications) of all staff and the centre's operating hours. This will give parents the opportunity to select the services that suit their situations.

Facilities

Commonly, regulations cover childcare facilities in great detail. They specify space per child, and require natural light, outdoor areas, child-specific toilets and wash basins and staff to observe children at all times. While these requirements are inspired by well-meaning considerations, a moment's reflection leads one to the realisation that most homes fall far short of these standards of appointment.

Therefore, would it not be appropriate to require such facilities as a pre-condition of people having and bringing up their own children? The logic of requiring such standards in childcare centres suggests that they should be extended to all facilities where children are under care. Indeed, they are all the more necessary where the child is living in premises for 24 hours per day, rather than the 40 hours per week in a centre.

The logic of current regulations could dictate that:

- children should not be brought up in tall apartment blocks, especially where there is little natural light;
- nobody should be allowed to have children, unless they have first completed the required tertiary qualification;
- there should be at least two people present to supervise children at all times, especially where there are more than three infants in the family; and
- homes in which children reside should be remodelled to ensure the wash basins and toilets are appropriate for those of small stature and those using the facilities can be observed at all times.

The childcare regulations represent a requirement that parents abandon their decision-making to a bureaucracy with regard to their children. The parents are the agents of their children. They make decisions for them. Incorporating a government bureaucracy into this process both undermines this and has adverse cost implications.

Enlisting the private sector to supply an increasing share of childcare has moderated the increase in costs over the years. But, adopting less prescriptive regulations can *lower* the cost of childcare and, hence, reduce the burden on working families.

Differing approaches to childcare

In summary, the approaches to the provision of childcare can follow one of two routes.

The first is having government specify in considerable detail:

- who may provide the service;
- the facilities in which it may be provided;
- the staff who may support the provider;
- the times that the facility may be open;
- the age profile of each facility's group of children;
- the ratio of staff to children and the nature of that ratio between staff of different credentialed expertise; and
- the prices at which the service should be provided.

The alternative approach is to insist on rigorous publication of information about the centre, so that parents can choose the quality of service they consider appropriate to their needs, preferences and resources.

Childcare is an important issue for parents. It is also a 'repeat purchase' and dissatisfaction will lead to a change in supplier. This places pressure on the centre to perform in order to gain and retain its customer base. With open access to the provision of childcare services, each centre must continuously strive to provide value. This market-based approach allows needs and offerings to be matched without the intercession of a bureaucracy.

The removal by the state of the need for parents to exercise control may also lead to a form of 'moral hazard'. Where a bureaucracy is vested with the responsibility to ensure a standard is maintained, it is likely that users will be less vigilant in undertaking these tasks themselves. Users will assume that others are undertaking the scrutiny role that they themselves would have performed previously. Fewer resources may be applied to the scrutiny as a result. It may also mean less effective scrutiny where regulators are focused on a rule.

The present regulatory tumescence is bringing high costs and denying capable carers without the skills or resources to pursue a university degree the possibility of offering the service. Parents should be allowed to make choices about the class of childcare they wish to afford for their children, just as they do about the nutritiousness of the food they serve or the quality of the housing they live in.

10A.4 Aged care

While children are a relatively declining business, the elderly form a growth industry. Various estimates put the number of Australians with 'profound or severe core activity restriction' at 800 000 in 1981, 1 210 000 in 2001, and growing to over 2 million thirty years from now.

Parents are the guardians and agents of their children, and children can make their views readily known to their parents. However, this same process is far less strong with the elderly. Many older people are less articulate than children are and the parental link with their own children is often weaker. Moreover, current policy approaches mean that there is a shortage of aged care places, compared to a demand-driven number of childcare places, where alternative providers need to operate on a highly competitive footing.

There are about 76 000 nursing home places and 64 000 hostel places. The present system followed a period of explosive growth with the introduction of

Commonwealth Government benefits in 1963. Since then, policy has been designed to hold back public expenditure. The Commonwealth subsidy to nursing homes is estimated at \$30 000 per bed year, accounting for roughly 80 per cent of costs. The Commonwealth also controls the number of subsidised beds and requires accreditation and compliance with building codes. State regulations usually specify staff ratios and other matters.

Most nursing homes are ‘for-profit’, with the remainder generally run by charitable bodies. The proportions are reversed for hostels, which remain free to negotiate prices.

The Aged Care Standards and Accreditation Agency replaced the Outcome Standards Monitoring Program. Inputs were, and continue to be, the main controls used. The major financial control is the Core Aggregated Module (CAM). This is a formula that determines the subsidy provided. The homes themselves have to provide funds for spending that is greater than 1.5 per cent above the subsidy level (they retain only the first 1 per cent of savings below the CAM). The inflexibilities of the system have been subject to a number of criticisms, as discussed by the PC (1999).

Even so, the current schemes remain inflexible, perhaps inevitably so. It is a confluence of the following factors.

- First, costs are largely covered by the Government with demand, therefore, well above a market-determined capacity to pay. Unlike childcare, the Government is not wedded to the idea of maximising the number of those being cared for.
- Second, resident contributions are fixed by a series of caps, the most important being a maximum contribution of 87.5 per cent of the single income pension. The Commonwealth Government has some asset testing in place, but this is estimated to bring in only 2.5 per cent of revenues, even by 2007. Some of the complexity emanated from the Government’s failed attempt to introduce greater user-pays in 1997.

The PC made several recommendations to increase the flexibility of the aged-care system. These included sensible tidying up approaches to a system that they envisaged would remain input controlled, such as:

- allowing people to pay for greater services and, thereby, allowing differentials and alternatives; and
- avoiding incentives to use up the entire subsidy available.

The PC suggested a continuing focus on input controls. After all, a system exists where not only are the costs basically covered by the Government, but inputs —

wages, facilities and treatment — are also controlled. Add to this the inability of many users to make their own decisions (or have a family member act as their agent) and the only option that remains is to devise a better set of benchmarks to manage the system.

One matter that was not directly addressed by the PC was the bed allocation system. At present, to rein in costs, new beds are licensed according to a formula based on the needs in each region. Applicants submit proposals and the Secretary of Department of Health and Family Services conducts a ‘beauty contest’ based on these applications. The allocation of the permits is valuable, with each allocated bed worth approximately \$27 000. Hence, a better approach from a public policy perspective would be to specify standards and allocate beds based on price or bid.

An approach canvassed by the Victorian Department of Human Services offers a path to inject greater efficiencies, by enhancing competitive tensions in the provision of aged care facilities. Victoria proposed a subsidy per person, with the recipient being able to purchase home-based or institutional-based care. This approach shows some promise by allowing consumer choice and competition to play a role.

That said, as the Australian Institute of Health and Welfare maintains, outcome-based measures are not all that useful with chronic care.

References

- AIHW (Australian Institute of Health and Welfare) 1999, *Australia’s Welfare 1999, Services and Assistance*, AIHW, Canberra.
- FACS (Department of Family and Community Services) 1999, *Child Care in Australia*, AusInfo, Canberra, July.
- PC (Productivity Commission) 1999, *Private Hospitals in Australia*, Commission Research Paper, AusInfo, Canberra.

10B The case against market-based regulation

Robert Fitzgerald AM

Community Services Commission

10B.1 Introduction

In a rapidly changing environment, the community services sector has been subject to some of the most profound changes in the last decade that it has ever faced. No less so has been the intense examination by Government of how, by whom and under what regulations community services should be delivered.

Before we can actually debate the point about whether or not market-based regulation is the best way to regulate community services, there are some fundamental characteristics about the community services sector that are important to be identified. All regulation must be examined within the context of the specific industry, its consumers and the outcomes sought. This is critically important in the area of community services.

10B.2 Features of community services

Often, it is easy to think of community services simply as being around the issues of childcare or aged care, which Alan Moran has done. However, community services comprise an extraordinarily broad range of services including:

- child protection;
- substitute care for children who have been removed from their families and are living away from home;
- support for families who are in crisis;
- accommodation and other services for people with disabilities;
- home support services for the frail aged and people with disabilities;
- services for homeless people;
- residential aged care;
- childcare; and
- drug, alcohol and other rehabilitation programs.

The Australian Institute of Health and Welfare (AIHW 1999) reported that in 1997-98, total expenditure on welfare services was estimated at \$10.9 billion or 1.9 per cent of Gross Domestic Product, not including social security cash payments. Interestingly, while 64 per cent was funded by the Government, 25 per cent was funded by users and the remaining 11 per cent was funded by the non-government agencies' own fund raising or other activities. More importantly, there are somewhere over 10 000 community welfare organisations, employing just over 320 000 people and volunteering the equivalent of 50 000 full time employees.

If we examine the forces that have impacted on the community services sector in recent times, we can see that, just like the general commercial world, the sector has been under considerable and sustained pressure to change. These pressures include:

- competition policy;
- the preference for market models even where the market has to be created;
- a redefining of Government to be only providers of core business;
- a desire to restrain or reduce public expenditure;
- shifting influences and ideologies in social policy;
- client based funding;
- managing for outcomes; and
- increasing use of 'for-profit' providers (ACOSS 1999).

But much more importantly, when we look at the issue of regulation, we need to understand some of the aspects of the consumers of community services. Regulation, as has been described today at this conference, should be about:

... a minimum necessary to achieve stated objectives and to avoid unnecessary restrictions. It should be targeted at the problem and not impose unnecessary burden on those affected.

In community services, the impacts that we have to examine are those impacts on the consumers as well as the providers. These consumers have a number of characteristics:

- the involuntary nature of most community services, in that, in many cases, people do not choose to use community services but may be forced to do so because of need or other circumstances beyond their control;
- the nature of community services places consumers in a position of dependence because of lack of alternatives should they be dissatisfied;

-
- age and capacity of consumers, particularly children, the elderly and people with disabilities, may limit their capacity to effectively advocate for themselves or to form effective consumer groups;
 - life circumstances of consumers may make it difficult to deal with more than their own immediate needs;
 - consumers often have a low sense of entitlement and a complete lack of knowledge about their rights; and
 - many consumers have few supportive personal relationships, resulting in social isolation and a lack of advocacy and support (Tang 1996).

These create potential risks for consumers and such risks are often exacerbated because of the failure of appropriate regulatory frameworks.

In this field of activity it is not correct to argue that regulation should only be applied where there is a net public benefit, but not where there are only benefits for certain groups of consumers. In the area of community services, the major need for regulatory requirements is to protect the safety and wellbeing of particularly vulnerable consumers, having regard to their particular needs, aspirations and circumstances. While consumer groups may be small relative to the overall population, their circumstances often require specific attention. This is not a ‘special interest’ pleading. It is simply understanding the nature of the consumers and their circumstances.

Today, we have been urged to the view that one should start with the premise of no regulation and then move through a series of graduated processes including self regulation, quasi-regulation (such as codes of practice) before embracing black letter regulation. In much of the community services sector, this approach has already been undertaken. If we look at areas such as nursing homes, we have moved from a situation of virtually no regulations and the extraordinary failures in terms of consumer protection, and the safety and wellbeing of consumers being compromised, to what would be called black letter regulation, compulsory standards and accreditation. Even Alan Moran acknowledges nursing homes as being different from childcare.

But similarly in childcare, often minimum standards of care and safety were not practised in previous times, thus a tighter regulatory environment emerged. This is also true in the areas of substitute care for children, the provision of homeless persons’ accommodation and so on.

Thus we have often seen that in community services markets (even when a market comprises both non-government and ‘for-profit’ organisations), the absence of adequate and appropriate regulation has often meant that the rights, entitlements and

safety of consumers have not been protected. Importantly, with such vulnerable client groups, the failure of such regulation has left people in dangerous and vulnerable environments. The vulnerability of consumers in this area is not simply about socioeconomic circumstance. Rather, vulnerability is often the result of physical or mental conditions, extreme emotional distress or pressure, or simply need arising out of circumstances beyond the control of the individuals.

10B.3 Objectives of community services

To determine whether or not regulation is appropriate, we have to determine what it is that we are seeking as outcomes from community services. Outcomes include:

- accessible, affordable and appropriate services;
- enhancement of people's emotional and physical wellbeing;
- the building of capacity within individuals, families and communities to be able to maximise the benefits of all activity and to sustain those benefits over time;
- the desire to encourage social and economic participation; and
- recognition and safeguarding of people's rights and entitlements.

If these are the objectives, what form of regulation best suits?

10B.4 The regulatory options

Contractual regulation

At the present time, we have to understand that there is not only overt or black letter regulation, but increasingly what we can call contractual regulation, which is much less transparent but just as prescriptive. There is a fiction in this debate in relation to whether or not regulation has increased. Often in community services, particularly with the increased contracting out of services, all that has happened is that regulation has grown markedly, not through legislation, but rather through contracts. Contractual regulation now governs much of the delivery of services in the community service system. If legislated regulations were further decreased in the statute books, they would quickly re-emerge in funding contracts but without the transparency or review processes provided through black letter regulation.

Indeed, as governments move out of service delivery, the level of regulation in various forms has increased. Most regulation is appropriately designed around the

issues of accountability, disclosure of information to consumers, and ultimately consumer protection and safety.

We have been asked to look at a couple of specific issues.

Liability law

One contentious issue is whether liability law should be sufficient, as distinct from the need for legislative regulation. Those who propose liability law as being sufficient should again understand the nature of the consumers within the community services system. The common law often requires extraordinarily costly and lengthy litigation for the principles of public liability to be established in particular areas. This we have seen in tobacco and alcohol cases. We will see this in the gambling area. Liability law also emerges after profound losses have been experienced. Too many suffer during the process of establishing what are publicly acceptable liabilities or obligations on the part of providers. More importantly, liability law rarely embraces the notion of best practice, but rather minimum standards. For those most marginalised and vulnerable, as users of community services often are, that is often not acceptable when we are seeking to maximise beneficial outcomes. It is correct that public liability law should provide a basis on which regulation should be built. Regulation builds on, and should not be in conflict with, the development of sound liability law. A society's expectations change over time and this is reflected in the common law. So too regulation needs to change. However, a mere reliance on liability law will do a great disservice to the safety, wellbeing and general rights of consumers of community services.

Service standards

That is not to say that there is not too much regulation which is often unnecessarily burdensome, and irrelevant to the stated objectives of regulation. It is not to say that in recent times we have not over-regulated various aspects of community services. Alan Moran, points to a number of such issues in relation to childcare. However the question remains who sets the standards, and I would strongly argue that consumers, particularly in the childcare and aged care, have increasingly demanded a higher quality of care and higher standards from service providers. More importantly, they want to be sure that such standards are being implemented across the system, that they do have a regulatory base and that accreditation ensures that the standards are actually being implemented. Is it wrong therefore for Governments to respond to the demands of consumers by seeking to enshrine these aspirations in law?

Outcome versus input regulation

Another issue is the notion of outcome versus input regulation. One of the greatest pressures on the community sector has been the move to outcome-based funding and performance measures. It is right that we should seek to have regulations that are more outcome based. There has been too much emphasis on inputs rather than outputs. Nevertheless, in the community services area it has been extremely difficult to actually define in clear terms the nature of outcomes. There are many difficulties with definitions, and some would argue the ability to actually define what is an appropriate outcome in many of these areas is impossible. Nevertheless, I support the move to greater outcome or output regulation but only when we can properly define and identify such outcomes or outputs in consistent and meaningful ways.

10B.5 The regulatory framework

Throughout the day we have heard many different approaches in assessing whether or not regulation should be applied to particular areas. One such approach has been to ask the question ‘will the market on its own produce a superior outcome than one obtained through regulation?’ Or put another way, ‘will the market produce the same or better outcome with regulation as it would have done without regulation?’

I am quite prepared to say that in most areas of community services, without appropriate regulation, then the outcomes in terms of consumer protection, safety and wellbeing would be poorer. The normal market pressures and forces do not exist in many areas of community services. The normal consumer pressures simply do not apply. True, that is less so in childcare but this is only one area of community services. Even in this area, consumer choice is restricted, especially for lower income clients. If we look toward what might be appropriate areas for regulatory intervention for consumers of community services, these could include:

- service standards which have a legislative base;
- monitoring mechanisms;
- the strengthening of consumers’ rights and the obligation of providers to inform consumers of these rights;
- mechanisms to protect and promote these rights;
- the right for consumers to be able to participate in decisions that affect them;
- appropriate complaint and appeal mechanisms; and
- the need for the disclosure of information, including provision of reasons for decisions.

It is the case that, until very recently, people with disabilities and people in aged care facilities did not have the right to participate in decisions which affected their own circumstances. They were not given information about their own care and often never consulted. For those of us who are articulate, well educated and confident, we can often demand our rights to knowledge, information and our right to participate in decisions that affect us. With resources, we can easily move to another provider. However, many consumers of community services do not have that same capacity. And even if they did, when they receive these services, mechanisms to enable them to redress decisions that adversely impact on them have often been missing.

I believe that the community services system is better served where we build on the public liability law through appropriate regulation. The regulatory framework should include:

- broad black letter regulation;
- contractual terms and conditions;
- standards in action; and
- accreditation.

Briefly, regulation should set out the general objectives and principles desired. It would identify the broad outcomes to be achieved and set the minimum standards and duty of care required of service providers. In particular, it would concentrate on disclosure of information, informed consent, involvement in decision-making processes, and mechanisms for redress.

Contractual or funding agreements would set specific outcomes (or inputs where appropriate). It would require the provision of certain standards and allow for evaluation and review of service provision.

Standards in action would actually set out appropriate standards to be applied. It would allow for the benchmarking between various services. At the present time benchmarking is almost impossible in many areas of community services because of the lack of usable data collection and the failure for standards to be applied in a consistent and orderly fashion across service providers.

Accreditation allows the consumers to purchase or access services with confidence. Of course, well informed consumers are able to find out information about a particular service and compare that with another. Yet there is an overwhelming sense that accreditation does provide a minimum baseline level of quality assurance. In the health area, we would certainly not accept the situation where hospitals reverted to an unaccredited environment or where standards were left simply to the

provider to pick and choose. Neither should we accept such a situation in other areas of community services.

After all this, we then come to the last question as to who should provide a given service? Often decisions in recent times have been driven by the notion that Government is a poor provider of services and that services should be contracted out in a competitively tendered manner. The decision about who should provide a service, whether there should be competitive tendering arrangements or negotiated contracts is really the last question after the framework to ensure quality and consumer protection has been put in place. At the end, market regulation is unlikely to, and has been demonstrated not to provide the highest quality of consumer protection for those who access community services generally. It is true that there can be inappropriate and over regulation. But the need for a regulatory framework, through legislation, contract, accreditation and approved standards seems to be best way forward to ensure the safety and wellbeing of users of community services in Australia.

References

- ACOSS (Australian Council of Social Services) 1999, *Common Causes, Relationships and Reforms in Community Services*, ACOSS Paper 102, November.
- AIHW (Australian Institute of Health and Welfare) 1999, *Australia's Welfare 1999, Services and Assistance*, AIHW, Canberra.
- Tang, A. 1996, *Contracting Out in Community Services*, Community Services Commission, Sydney.

Response — *Alan Moran*

Institute of Public Affairs

There is a spectrum or a continuum of issues regarding community services — ranging from childcare, through to nursing homes and child protection. These issues highlight the difficulties in applying non-regulatory market solutions to achieve the efficiencies that society desires elsewhere in the economy. Yes, volunteers are valuable, and it is important to use non-market instruments to knit communities together.

In most of the points Robert Fitzgerald made, there is little to disagree with. By and large it is an application of the principle that society should get the best possible value for money. Robert Fitzgerald talked about community services needing to be accessible, affordable and appropriate. This forms a good basis for assessment, but what does ‘appropriate’ mean? All too often, appropriate means something better than affordable, and this can lead us into a wilderness of excessive cost.

There are excessive costs built into childcare, driven by the users perhaps, as well as the bureaucracy. But when the taxpayer is paying 60 per cent of the bill, there is a very strong reason for people to want something better than they would normally have. Robert Fitzgerald raised the point that individuals demand higher standards in a hospital than they do in their own home. But individuals typically do not seek serious medical treatment in their own home. Childcare, on the other hand, provides care that would otherwise have been provided by the parent. It costs more than it should and does not receive much scrutiny. It could be said that Government has responded to consumer or electoral needs, but we ought to be able to say that this is not an appropriate use of government funding. This is difficult in the political arena. For example, the middle class derailed the Government’s attempt in 1997-98 to increase the contribution of users to nursing home care, so the community as a whole continues to contribute to the retirement of the elderly.

There are some choices to be made. We must try and find the most affordable way of doing things, compatible with the standards and sensitivities of the community. Community services constitute 1.92 per cent of GDP and are likely to grow rapidly, especially in the area of aged care. We must find ways to economise on public funds. This should not involve giving people whatever we think appropriate at any cost. In the case of childcare, there is potential for cost savings that can be reallocated to other areas of the community. More outsourcing in areas like child protection will also save money.

When unnecessarily high standards are imposed, some individuals also miss the opportunity to apply their skills.

Response — *Robert Fitzgerald AM*

Community Services Commission

The word ‘appropriate’ is in fact a commonly used word, particularly in the area of social and affordable housing. It means providing a service that is appropriate to the needs of the individual. This is a major issue, because often it does not happen.

Alan Moran is correct in identifying the issue of who judges whether regulations are providing an unduly high level of standards. But if standards are lowered, it is always the same group who loses — those who could not afford higher standards otherwise. The residualisation of services in the community service system is a harsh reality of history and current times. Residual services led to lower quality nursing homes, boarding houses and childcare facilities. They led to inappropriate, abusive behaviour. They led to people’s rights being denied. In some areas, such as childcare, the current standards may be inappropriate. However, it must be acknowledged that if standards and subsequently government contributions are reduced, there may be losers in the community. This group could be large.

The downsizing of community services has had political consequences. Unless appropriate standards are provided, and the Government is prepared to support such standards, then the community itself (not just the consumers of the service) will demand certain service standards. Business always attacks middle class welfare, but always supports it. Middle-class welfare has risen very noticeably in Australia, particularly over the past four or five years, and is unlikely to fall. But Eva Cox has said that if welfare is not provided to the middle class, they will deny the resources to provide it to those in genuine need. This is a contestable point, but one that everyone should be aware of politically.

General discussion

The discussion focused on:

- the role and type of regulation;
- information availability;
- nursing home regulation; and
- childcare regulation.

The role and type of regulation

It was noted that Robert Fitzgerald focused specifically on the vulnerable members of the community. If there was an alternative way to deal with the vulnerable, the need for government regulation and control would be removed and a market-based approach could be adopted.

Robert Fitzgerald clarified that vulnerability was a broad concept, covering not only those who were economically or socially disadvantaged, but also those who were vulnerable because of health or circumstance. He also noted that regulations to achieve standards were widely accepted for many goods and services, though not all specified, as for community services, who could buy or sell the good. He conceded that there were alternative ways of approaching vulnerability other than via regulation. Issues of education, empowerment and resourcing were relevant. As the power of the individual to make informed choices and to exercise that choice increased, the need for prescriptive legislation was reduced in some areas.

Robert Fitzgerald had talked of three levels of regulation to protect the vulnerable — minimum standards, contractual arrangements and accreditation. A participant argued that in this regime, Government was still supplying the services, though by proxy. With a proper purchaser/provider split, with Government as funder, and with purchasing agencies to help the vulnerable, it may be possible to work towards the market model proposed by Alan Moran.

Robert Fitzgerald thought it debatable whether a better purchaser/provider split would deliver better outcomes. He doubted that it would require less regulation to support minimum standards, but noted that it could reduce regulation to do with choice and informed decision making.

Alan Moran observed that many goods and services were over-regulated, often to protect the incumbent producer from competition. This sort of regulation had adverse implications for an efficient, productive and dynamic society.

Information availability

A participant supported Alan Moran's conclusion about making available to consumers information on the comparative performance of different institutions. In this context, it was noted that several attempts had been made to gain inspection reports on nursing homes from the relevant Commonwealth departments in a form that could usefully be published and used by consumers to make choices. The relative lack of success was attributed to industry capture. The reforms in this area to date had been limited.

Nursing home regulation

Under the old regulatory regime that applied until 1997, Standard 1.1 stated that all residents were able to use the doctor of their choice in the nursing home. This was an important standard that addressed the difficulty of infirmed consumers voting with their feet. They instead had the option of using their family doctor to act as their proxy. Alan Moran was asked whether he favoured a reinstatement of the right to bring in the family doctor.

Alan Moran did not necessarily favour a review of a nursing home resident's right to choose their doctor, because of the potential for medical practitioners to then exercise monopoly power. There were examples in the United States of groups of hospitals and nursing homes successfully challenging the monopoly of medical practitioners and driving down excessive costs. Individuals were not able to make sourcing decisions easily themselves, but their agent — their HMO, their health fund or, perhaps, their nursing home — could make decisions and thereby reduce costs, for the benefit of all.

Robert Fitzgerald noted that permitting agents to make sourcing decisions required a regulatory framework to ensure that the service provider made information available to the advocate on behalf of the client. A move to an outcomes-oriented system was appropriate, but it still required regulation to ensure the outcome. In the area of disability, for example, Australia had moved from tight regulations to broader principles, but these were still needed. It was naive to believe that achieving the desired outcomes could be achieved in a regulation-free environment.

Childcare regulation

It was suggested that Alan Moran's approach to childcare was a little simplistic in attributing all cost increases to regulation. Other changes in the childcare sector had contributed to significant cost increases, particularly wage increase that were not

necessarily attributable to increased educational qualifications. There were also multiple government influences in childcare, including local, State, Territory and Commonwealth Government regulation. Finally, there are supply shortages in some sectors, even in family day care.

Alan Moran responded that the costs of childcare were higher than expected, considering its resources were commonly available in society.

There was concern that the societal benefits from childcare were being ignored. Some overseas studies were beginning to show evidence of such benefits, particularly in the area of early childhood development. Childcare and other services for three- to four-year-olds were delivering significant returns on investment. These returns from childcare needed to be considered along with the input costs.

Alan Moran responded that the evidence regarding societal benefits from early childcare had varied over time. Both Alan Moran and Robert Fitzgerald agreed that if there were societal benefits, these should be acknowledged.

11 Regulating gambling: a ‘market friendly’ approach to the social impacts

Gary Banks[†]

Productivity Commission

11.1 Introduction

In February 2000, newspapers around the country highlighted the story of a 19-month baby boy who had been left unattended in a car in sweltering temperatures while his mother went to play the pokies in a Melbourne hotel. By the time he was rescued, the child had sustained severe brain damage and he subsequently died. Two months later, in April, a struggling pensioner from near Nowra in NSW won the \$1.15 million first division prize in Oz Lotto — news that also attracted considerable media attention. These two contrasting incidents exemplify the ambivalent character of gambling for Australians. Mirroring this complex picture, the Productivity Commission in its recent inquiry report, *Australia’s Gambling Industries* (PC 1999), found that gambling produces substantial benefits to Australians, but that these industries also occasion large social costs.

The two-sided nature of gambling means that policy should aim to preserve the capacity for millions of Australians to derive pleasure from gambling, while reducing the potential for harm. Achieving such a balance requires good information about the impacts of gambling activities and scrupulous attention to regulatory process and design.

In practice, the rapid liberalisation and expansion of gambling industries in Australia has had the benefit of neither. Gambling has become enmeshed by an incoherent set of regulations that reflect historical accident, expedience and counteracting objectives — with government revenue predominant — to which

[†] This paper draws heavily on the Productivity Commission report *Australia’s Gambling Industries* (PC 1999). Additional commentary is the responsibility of the author. The assistance of Dr Ralph Lattimore is gratefully acknowledged.

multiple jurisdictions have added a further layer of complexity. But until very recently, most governments have failed to grapple seriously with the social ramifications.

Regulation reaches into almost all aspects of the conduct of gambling operators (figure 11.1).

A striking feature, in this era of competition policy, is the widespread restrictions on entry to the gambling industries. These include the common practice of establishing exclusive rights to gambling (such as State lotteries and Totalizator Agency Boards (TABs¹), and sometimes casinos), curbs on gambling machines in particular locations (such as shopping centres), and more favourable entry conditions for certain forms of venues (such as mutual clubs versus for-profit hotels).

Even if an enterprise is permitted entry, it is likely to face a range of quantity restrictions — such as limits on machine or table numbers, or prescribed hours of operation. The restrictions can create economic rents that are then subject to expropriation through government taxes and licence fees. At the same time, however, governments also typically prescribe maximum prices (and sometimes minimum ones).

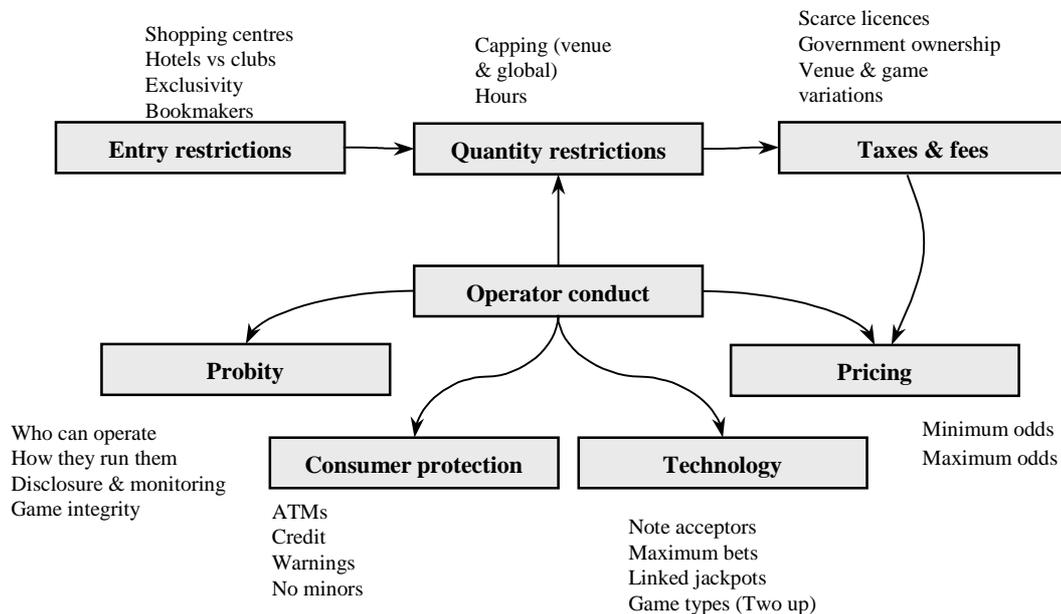
In the interests of probity, governments enforce who can operate a given gambling venue — particularly casinos — and how they do so. But the strictness of such probity regimes varies among different types of venues for reasons that are unclear.

Governments also determine standards for gaming products, such as gaming machines, including the testing of any chips that determine player return rates. They also prescribe, and in some instances prohibit, gambling technologies. It is legal to bet on thoroughbred horses and greyhounds, but not to arrange bets on Arabian or quarter horses. Two-up is legal only some of the time. Machines have restrictions on betting limits, jackpots and, in some cases, note acceptors.

Governments have also paid some attention to consumer protection and problem gambling and, indeed, some of the regulations reviewed above are intended to protect consumers. However, until recent regulatory changes, the measures have generally been poorly targeted and relatively weak, and have occupied a minor role in the function of regulatory agencies.

¹ Totalizator Agency Boards control most off-course betting on horse, harness and greyhound racing within particular States in Australia. All money bet on a race goes into a pool and, after the completion of the race, the pool (less a small percentage) is shared between those holding winning tickets. Unlike with bookmakers, the odds of winning are not fixed and are only determined at the start of the race. This is also known as *pari mutuel* betting.

Figure 11.1 The nature of gambling regulations



Looking through the lens of good regulatory practice, regulatory arrangements for gambling are not only pervasive and anticompetitive, but also characterised by other significant flaws. In particular:

- it is often hard to find explicit statements about what a regulation is intended to achieve or why that is a sensible goal — the starting point for good regulation;
- regulators have rarely sought authoritative assessments of the impacts of changes to gambling regulation; and
- some regulations, as a result, have questionable effectiveness in achieving important goals.

This paper looks at each of these aspects in turn before considering some institutional prerequisites for regulating gambling in the public interest.

11.2 Assessing rationales for gambling regulation

There appear to be four broad rationales that have shaped gambling policy and regulation throughout Australia:

- generating government revenue;
- economic development;

-
- helping special industries; and
 - social and consumer issues.

Generating government revenue

Against a background of States' increasing dependence on Commonwealth revenue and a perceived lack of alternative State taxes, governments have seen gambling as an attractive source of revenue from taxation and licence fees. Arguably much regulation aims to facilitate revenue maximisation by creating captive markets and monopoly rents. As noted in Tattersalls' submission to the inquiry:

Given gambling's effectiveness as a revenue raising device, governments have not long been able to maintain a policy of prohibition and non-involvement. The potential for revenue loss to other jurisdictions has also been a strong factor encouraging the spread of gambling services. (Tattersalls 1999, p. v)

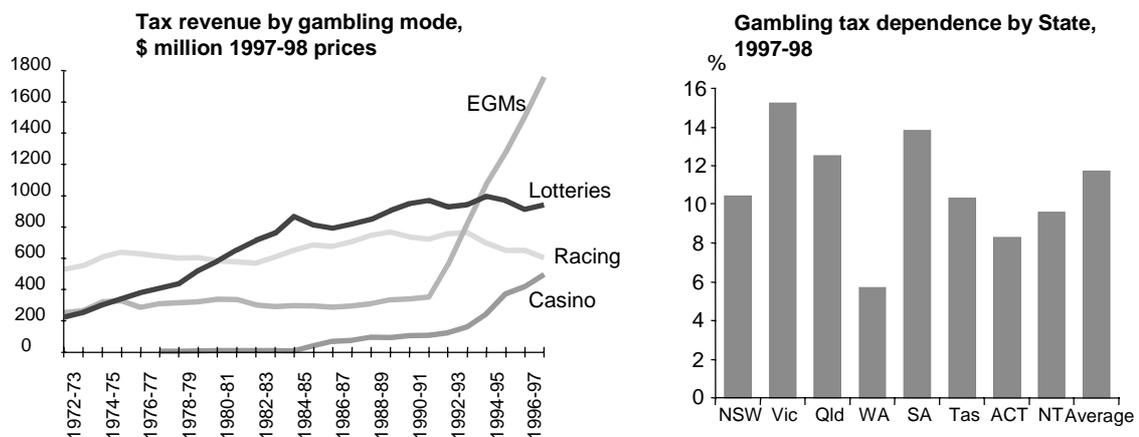
The earmarking by governments of a proportion of gambling revenue for 'good causes', together with the ambivalent attitude by the public to gambling, have helped to further legitimise high taxes. The States' distorted incentives to use gambling as a revenue raiser have also been compounded by the distribution methodology of the Commonwealth Grants Commission, which penalises below average tax raising efforts.

Liberalisation of gambling — particularly electronic gaming machines (EGMs) — has led to substantial rises in tax revenue (figure 11.2). Most States, with the exception of Western Australia, now derive a sizeable share of their revenue from gambling.

Revenue generation has been a dominant influence on the nature and form of regulation. However, the need to generate tax revenue should be balanced by consideration of any adverse effects on consumers, producers and others. Gambling taxes, like other taxes, need to be evaluated on efficiency, equity and social grounds. The analysis, as for other aspects of gambling activities, is not straightforward.

It is well known that even relatively high taxes can be efficient if demand for the product is unresponsive to resulting price increases. The evidence, while equivocal, does suggest relatively insensitive demand. In these circumstances, for most gambling forms, there would not necessarily be a payoff to efficiency from significantly reducing tax rates. Lotteries may be an exception, but their tax rates are already so high as to possibly outweigh the effects of inelastic demand.

Figure 11.2 The importance of gambling taxes



Source: ABS Taxation Revenue, Cat. no. 5506.0, various issues and Commission estimates.

Normally there are grounds for taxing more heavily those activities with undesirable side effects, to reduce their production or consumption. But in the case of gambling, the adverse side effects stem from excessive *spending*. And it seems clear that, irrespective of any doubts about recreational gamblers, most problem gamblers are unlikely to be sensitive to price changes. So higher taxes generally make for greater financial burdens on existing problem gamblers. However, not enough is known about the behaviour of problem gamblers to predict the outcome for them from lower taxes. Also, lower taxes could serve to encourage increased gambling activity by people who are at risk of becoming problem gamblers. This and other complications led the Productivity Commission to conclude that taxation is too blunt an instrument for reducing the social costs of gambling.

Gambling taxation is regressive, with lower income groups generally spending proportionately more on gambling — and thus shouldering more of the burden. The Productivity Commission’s analysis suggests that taxes on lotteries and gaming machines are the main sources of this regressivity. The equity issues are heightened by the unevenness of the tax burden among the poorest households, with some paying much higher proportions of their income in gambling taxes than others. However, equity outcomes from reducing gambling taxes would also depend on what alternative taxes were available to States and Territories to replace lost revenue, and their degree of regressivity.

In summary, there are both efficiency and equity grounds for experimenting with lower lottery taxes. While the levels of other gambling taxes are unlikely to be optimal, on the basis of available information there is not a strong, or unambiguous, case for general reductions.

But there are other, broader grounds for concern over the way revenue generation influences gambling regulation. Specific taxes on gambling are at least visible and open to change, as government sees fit. Fees extracted as a *quid pro quo* for a long-term exclusive licence — as in the Victorian gaming machine duopoly or Casino licences generally — are less transparent and flexible, and can weaken or eliminate competition.

Gambling-led economic development?

Advocates for the expansion of the gambling industries typically point to economic benefits in terms of additional expenditure, incomes, jobs and trade. But these ‘production-side’ benefits from liberalising gambling are often greatly overstated. In reality, they are modest compared with the economic benefits derived by consumers.

If the gambling industries had not been permitted to expand, the money spent on gambling would have been spent elsewhere. And most of the resources that went into the gambling industries would have been employed in other uses, creating similar levels of income and jobs to gambling itself. For example, the skills required of personnel in gambling venues are very similar to those in most entertainment and hospitality industries.

Thus, while there may be instances where additional jobs or income may have been generated — say in depressed regions — most of the resources in the gambling industries will have been diverted from other industries. The vocal opposition of retail traders to the expansion of gambling outlets is a visible sign of this underlying economic reality. By the same logic, however, that diversion should not in itself be of concern to policymakers, unless it *reduces* the efficient use of economic resources, rather than simply reshuffling them.

Some indication of the net economywide contribution of the gambling industries can really only be gained with the aid of quantitative economic models, notwithstanding their particular limitations in dealing with the social costs of gambling. Such economywide modelling was conducted by the industry, as well as by consultants commissioned by the Productivity Commission. The Commission’s analysis of these various studies, taking into consideration their different methodologies and assumptions, supports the qualitative reasoning about the industries’ likely net contribution to the economy. In short, the modelling indicates that changes in the size of the industry would have little impact on Australia’s GDP, consumption levels or labour market outcomes over the long term.

The real net contribution of the gambling industries, thus, depends on the extent to which consumers are better off through any enjoyment they obtain from gambling. The perceived production side benefits do not constitute a convincing rationale for gambling policy, nor do they provide grounds for regulators to promote the industry. That is the province of the operators themselves.

Helping 'special' industries?

A further apparent rationale for gambling regulation and taxation has been governments' desire to assist 'special' groups, such as the mutual non-profit club sector, hotels and the racing industry. The outcomes have included exclusive access to gambling, favourable tax rates and other concessional treatment. None of these provide a compelling case for regulation.

- The racing industry does provide benefits — in the form of an event that can be wagered on — to gambling operators. However, there are other ways of funding the racing industry that do not rely on anticompetitive arrangements, such as a single gambling provider.
- In some States, the club sector has privileged access to gaming machines, as well as lower tax rates on revenue. While the club sector plays a major role in Australian communities, the grounds for providing gambling to them on a preferential basis are weak. It begs the question of why gambling should be singled out in this way: why not preferential treatment for other goods and services, like beer or cinemas? If clubs serve a social role which governments believe requires support from taxpayer's funds, then transparent payments would generally be more appropriate than the allocation of preferential rights.
- The argument for liberalising gambling in order to revive the hotel industry also lacks a firm basis — on grounds similar to those applying to clubs.

A particular danger of the concern for 'special' groups in framing gambling regulatory policy is that it can divert the focus of regulators from their appropriate clients — consumers and the general public — to stakeholders with economic incentives to tilt outcomes in their own favour, to the detriment of the public interest.

Consumer and social issues

A far more compelling basis for regulation is to deal with general consumer and public interest concerns about gambling. The Productivity Commission concluded that the key rationales that should guide government regulation were:

- amelioration of the social costs of gambling;

-
- the need to ensure that consumers are adequately informed; and
 - probity controls — both to protect consumers and to reduce potential criminal activity.

There are three caveats that have arisen concerning the need for regulated consumer protection in gambling.

The first is that the existence of grounds for government intervention is not sufficient for regulation. Each regulation based on apparent consumer interest grounds has to effectively and efficiently achieve its goals, such that there is a net improvement. Some regulatory measures, whose putative goal is the consumer interest, do not achieve that goal. For example, reductions in accessibility may reduce the social costs of gambling, but with the exception of casinos, exclusivity arrangements for TABs or gaming machine owners do not reduce accessibility. They simply reduce competition. Thus, regulators need to scrutinise whether a measure ostensibly aiming to reduce adverse social impacts really does so.

The other two caveats about consumer protection as a rationale for gambling regulation have less foundation.

First, it is sometimes claimed that there are ‘gamblers with problems’, but not problem gamblers as such. The inference is that the causality from gambling to problems is suspect and that the people reporting problems with their gambling would have experienced problems in the absence of access to gambling. However, the evidence canvassed by the Productivity Commission on this point suggests that, while there are some people who would have experienced problems in the absence of gambling, for most, gambling creates or exacerbates problems.

Second, some advocates for the gambling industries have claimed that problem gamblers are ‘rational addicts’ (ACIL Consulting 1999, pp. 91–6). The hypothesis of rational addiction is theoretically sophisticated, but its essential proposition is that addictive behaviour is rational and forward looking. This implies that, at any point in time, an addict calculates that the costs of giving up addiction outweigh the costs of being addicted. The implication of the theory, if true, would be that many of the apparent costs of problem gambling (such as guilt, depression or conflict) were internal to the gambler — and thus, unlike true externalities, not a matter for public policy.

However, while interesting as a theory, it has limitations for understanding problem gambling:

- for one thing, the theory has not been tested for problem gambling;

-
- clinical psychologists who have conducted empirical studies see problem gambling as a behaviour in which players have weakened self-control or lose the ability to think rationally about their gambling actions;
 - it does not conform with the self-described experiences of problem gamblers, who regret the amount of time and money they end up spending; and, most tellingly,
 - it does not seem consistent with the attempts by problem gamblers to seek self-exclusion, or other pre-commitment mechanisms that bind what they recognise will be future uncontrolled behaviour — pre-commitment is the ‘smoking gun’ of unstable preferences (O’Donoghue and Rabin 1999).

Thus, while many of the costs of problem gambling may be internal, they are not, in economists’ parlance, ‘internalised’ in the gambler’s decision-making — and thus do have relevance for policymakers.

That said, the alternative to rational addiction is not complete irrationality. Problem gamblers can be ‘boundedly rational’ — periodically able to exercise control and anticipate the consequences of their actions. For example, they may take a limited amount of money to a venue to try to limit their losses. It is the scope for periodic rationality that offers hope for more effective harm minimisation, such as the pre-commitment measures discussed in section 11.5

11.3 Assessing the impacts of gambling

The proliferation of gambling opportunities and the growth in the gambling industries over the last 10 to 15 years have been remarkable. Real per adult spending on gambling doubled between 1987-88 to 1997-98, with average annual growth rates in the 1990s of around 8 per cent, compared to 1.5 per cent in the two preceding decades. Deregulation and technological developments have driven this increase in spending. These have been manifested in:

- a proliferation of gambling forms, which commenced with the spread of casinos and then of electronic poker machines, with lottery products also becoming more diverse and sports betting expanding (including through the internet);
- increasing accessibility and ‘convenience’ of gambling, which in most jurisdictions is now part of the suburban scene;
- a more rapid ‘tempo’ of gambling, through electronic machines with much higher spending rates than the old ‘one arm bandits’, as well as more frequent race meetings and lottery draws;

-
- privatisation of the traditional government-run gambling forms — TABs and lotteries — with involvement of large corporations, and increasing concentration of ownership in some areas; and
 - more pervasive advertising and promotion of gambling (including the use of gambling as a marketing tool for other products).

Despite the two-sided nature of gambling, this expansion occurred without a balanced assessment of its impacts within most jurisdictions. In particular, while there have sometimes been upbeat assessments of alleged multiplier effects of expanded gambling, there has been little disinterested assessment of either the benefits or the costs. One of the assigned tasks of the Productivity Commission's national inquiry was to take stock of the cumulative effects, both beneficial and adverse, of this liberalisation.

As previously indicated, the Productivity Commission principally saw the benefits as entertainment value to consumers, with over 80 per cent of Australians gambling in the last year. The common misconception that gambling generates no worthwhile benefits (apparently held even by Nobel-prize winning economist Paul Samuelson) can be attributed either to moral 'blinkers' or to the 'materialist illusion' that only tangible goods or services yield economic gain. The latter misses the point that people can derive pleasure or satisfaction from some activities regardless of any tangible output. Thus, as the Commission emphasised in its report, many people gamble because of the enjoyment they get from the venue, the social interaction, the risk, the thrill of anticipation or some combination of all of these. Gambling venues, such as casinos and clubs, can also provide an accessible, comfortable and safe social environment, which many people — particularly women, elderly people and ethnic communities — have found appealing.

The Commission's inquiry also found evidence of adverse social impacts, with an estimated 2.1 per cent of adults being categorised as having significant problems associated with their gambling (box 11.1). Expressed as a proportion of regular gamblers on non-lottery modes of gambling, the incidence of problem gambling rises to around 15 per cent. Perhaps more telling still for policy purposes, this relatively small sub-set of gamblers accounted for around one third of Australia's total expenditure on gambling, reflecting the relatively large annual losses incurred (\$12 000 on average, compared to \$600 for other gamblers).

Problem gamblers were unequally distributed among gambling modes. Gaming machines, racing and casino table games accounted for most difficulties experienced by gamblers (figure 11.3), whereas lotteries and scratchies posed negligible risks. Spending shares also varied markedly, with problem gamblers

accounting for an estimated 42 per cent of total spending on gaming machines compared to 5 per cent of lottery spending.

Box 11.1 What is problem gambling?

There are many definitions of problem gambling, but they share much common ground. Characteristics include control problems, guilt, chasing losses, interpersonal conflicts over gambling, and in some cases behaviours that affect society generally, such as stealing and absenteeism. Implicitly, problem gambling involves harm to either the gambler, those close to them or society.

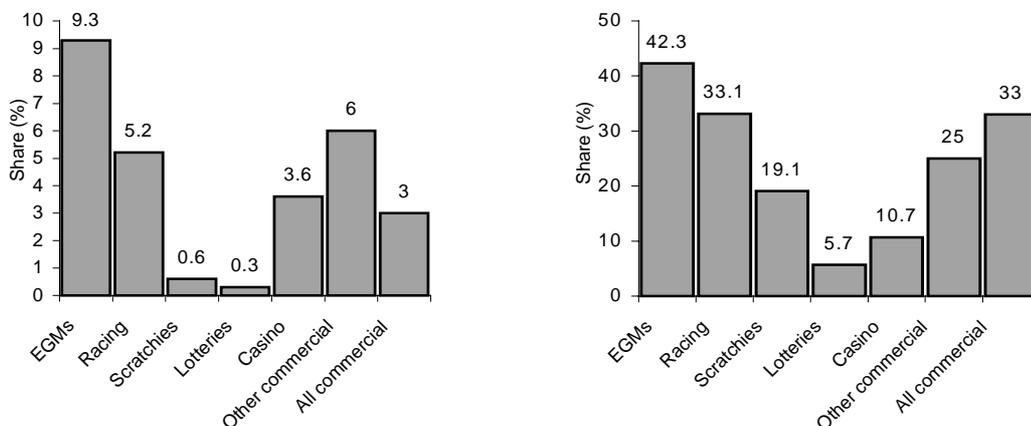
There remain disagreements about the frameworks for understanding problem gambling. Australian researchers have typically avoided the US ‘medicalised’ framework, in which problem gambling is seen as a psychiatric disorder. Rather, problem gambling is more typically conceptualised in a broader public health context, involving interactions between the gambling environment (such as game features and accessibility) and the personal characteristics of the gambler.

Moreover, problem gambling is best seen as lying on a continuum of severity. The Productivity Commission excluded minor problems from its policy analysis of problem gambling and only looked at those with moderate problems (such as control problems, guilt and lying) and the minority with severe problems (depression, suicidal intentions, stealing) associated with their gambling.

In doing so, the Commission used a much higher threshold for defining and screening problem gambling than is typically used in the United States.

Figure 11.3 Prevalence and spending shares of problem gamblers

Problem gambling prevalence by favourite mode of gambling Share of expenditure accounted for by problem gamblers, by mode



Data source: Productivity Commission, *National Gambling Survey 1999*.

The adverse impacts of problem gambling are not confined to the gamblers, but extend to family members, employers and others (eg through larceny and theft). The evidence suggests that 5 to 10 people can be affected to varying degrees by the behaviour of a problem gambler. In addition, there are demands on community resources and public services.

An attempt at quantification

Valuing all these benefits and costs is a hazardous task, given the lack of information about key aspects. Attempting to estimate the costs is especially problematic, as many involve intangible impacts on individuals. But the psychic or emotional impacts on problem gamblers and their families are costs for which a value should be assigned, in the same way that the pleasure or entertainment from gambling has a value. The difference is that only the latter value is expressed through actual market prices — proxy values have to be found for the former.

Therefore, in responding to its terms of reference, and because certain estimates by participants and others were being used in public debate, the Productivity Commission attempted to quantify as many of the benefits and costs as possible, to help inform judgments about what the net impacts could be. That said, the range of estimated values for both the benefits and the costs is necessarily wide, given the uncertainties involved.

The net outcome, deducting estimated costs of problem gambling from net consumer benefits (including tax transfers), ranges in aggregate from a net loss of \$1.2 billion to a net benefit of \$4.3 billion for 1997-98. Box 11.2 provides a brief explanation of how this was done. The estimates still leave out some potentially significant sources of cost that the Productivity Commission was unable to quantify, including gambling-related suicides and potential community impacts unrelated to problem gambling.

There are significant differences by gambling mode, with lotteries showing a clear net benefit, whereas gaming machines and wagering include the possibility of a net loss (table 11.1).

Box 11.2 'Ballpark' estimates of the benefits and costs of gambling

There are two dimensions to the calculations: the first involves estimating net benefits to consumers; the second involves estimating the costs of problem gambling.

The consumer benefits

Economists measure consumer benefits as the extra value that consumers derive from a product — in this case, gambling — above what it costs. This is known as 'consumer surplus'. Estimates were based on current consumption levels, with the most critical assumptions being about:

- the sensitivity of gamblers to changes in the 'price' — information is very poor, so that a range of plausible estimates were used; and
- the value that problem gamblers place on their (excessive) consumption of gambling; where it was assumed that they would spend on average an amount equal to that of a regular recreational gambler and get similar satisfaction levels (a generous assumption compared to US studies).

On this basis, the estimates of net benefits from consumption (including tax revenue) ranged from \$4.4 billion to \$6.1 billion a year for 1997-98.

The costs of problem gambling

The surveys asked respondents about a range of impacts from their gambling. Using this and other information about impacts, and drawing on various sources in valuing them, the Productivity Commission came up with the following cost ranges:

<i>Impact</i>	<i>Low (\$m)</i>	<i>High (\$m)</i>
Bankruptcy	1.3	1.3
Productivity loss	28	200
Job change	59	59
Police, court and jail	14	14
Distress of family and parents	756	2 933
Breakup, divorce and separation	417	1 120
Violence	2.8	8.3
Depression and suicide	502	1 230
Gambling counselling services	20	20

In total, these costs range from \$1.8 billion to \$5.6 billion. The wide range reflects the difficulty of putting dollar values on the intangible but important emotional impacts. (The methodology used, and differences from the estimates in the draft report, are explained in appendix J of *Australia's Gambling Industries*).

The net outcome

Deducting the estimated costs of gambling from the net consumer benefit numbers yields a range from a net social cost of \$1.2 billion to a net benefit of \$4.3 billion for 1997-98.

Source: See chapters 9 and 11 and appendices C and J of *Australia's Gambling Industries*.

Table 11.1 Measured consumer benefits, social costs and net impacts of gambling, by mode of gambling

\$ million (1997-98)

	<i>Net consumer benefit</i>	<i>Social costs of gambling</i>	<i>Net benefit^a</i>
Wagering	629 — 885	267 — 830	(201) — 617
Lotteries	1 232 — 1 498	34 — 106	1 126 — 1 464
Scratchies	219 — 266	24 — 74	145 — 243
Gaming machines	1 617 — 2 491	1 369 — 4 250	(2 634) — 1 122
Casino gaming	581 — 771	48 — 150	431 — 723
Other	103 — 184	57 — 176	(73) — 127
All gambling	4 365 — 6 076	1 800 — 5 586	(1 221) — 4 277

^a Figures in brackets represent a loss.

Source: Productivity Commission estimates.

Lack of knowledge about the probability of different net benefit outcomes across these ranges, or of the profile of costs and benefits for different levels of gambling, limits the policy usefulness of such calculations. Nevertheless, the potential magnitude of social costs associated with gambling are so large, particularly for wagering and gaming machines, that governments should seek measures that can reduce them. On the other hand, precipitate action that impacted adversely on the sizeable consumer benefits would be misplaced. Accordingly, regulations aiming to protect consumers and prevent or ameliorate problem gambling should be targeted in a way that does not curtail significantly the recreational value of gambling.

11.4 Capping: a closer look at a key regulatory response

A variety of evidence — from the Productivity Commission’s surveys, from counsellors’ databases and from clinical evidence — suggests that gaming machines are the prime source of problem gambling. Given persuasive evidence that accessibility of gambling influences the extent of problem gambling, it might seem that an easy way of moderating the adverse impacts of problem gambling is to have limits on machine numbers. Indeed, this has been a key regulatory response by governments to community concerns. In practice, it illustrates the need for careful assessment of measures intended to ameliorate problems if they are to be effective and avoid unintended (or even perverse) consequences.

There are caps on the number of machines — at a venue level or jurisdictional level or both — in all States and Territories.

- For example, Victoria currently has a Statewide cap of 27 500 machines (excluding Crown Casino), with hotels and clubs limited to 105 machines each.

Concern about the social impacts of gambling undermined the momentum that was gathering to increase the cap. The new Victorian Government has also mooted regional caps (Pandazopoulos 2000).

- The New South Wales Government announced a Statewide cap of 100 000 machines in early 2000, supplementing venue caps that applied to hotels and the casino (but not to clubs).

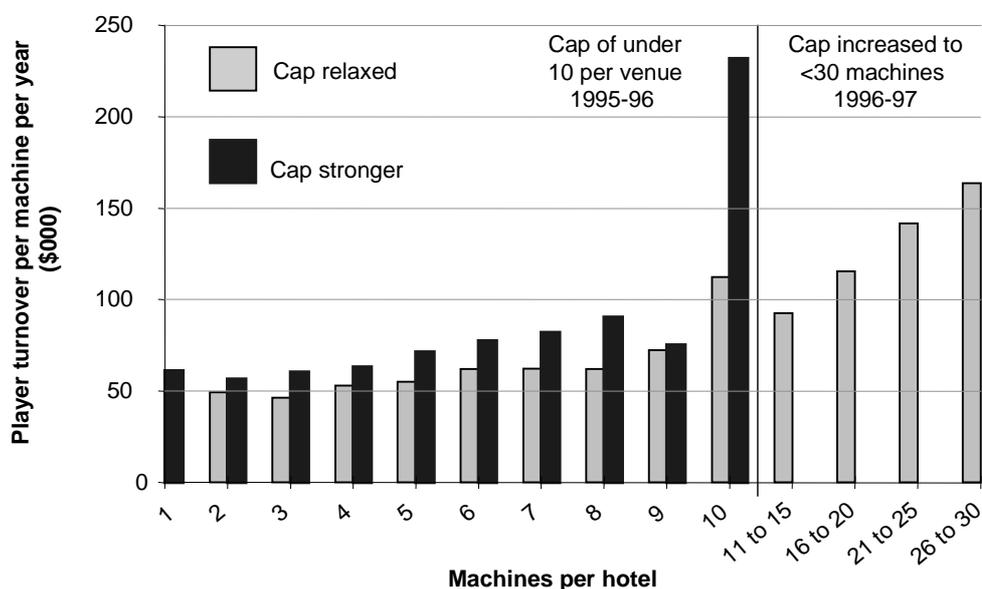
Both the concept of caps, and the levels at which they are set, are contentious issues. In the Productivity Commission's view, supply restrictions can only be justified to the extent that they can reduce social costs sufficiently to warrant any adverse impacts on recreational consumers.

Assessing the impact of caps is complex, as it depends not only on how tightly they are 'binding' (demand exceeding supply), but also on other aspects of the regulatory environment (such as price controls and governance) and on the way consumers and venues respond to constraints.

For one thing, once demand pressures mount, there will be incentives on operators and gamblers for the more intensive use of machines, which could exacerbate problem gambling. Thus, machine intensity is much greater in Victoria, where there are Statewide caps, than in New South Wales. It is also notable that when hotel caps were relaxed in New South Wales, turnover per machine in those hotels that were most constrained by the cap fell significantly (figure 11.4). Thus, even though there were 34 per cent fewer machines when the cap was binding, total turnover was only 2.5 per cent less. Looked at the other way, reimposing lower caps would, on this experience, have only a marginal effect on player losses.

Moreover, to the extent that venues can raise 'prices' (reduce the odds or payout ratios) in response to demand pressure on scarce machines, this is also likely to increase the spending of existing problem gamblers. To the extent that venues cannot raise prices, sheer congestion and queuing might somewhat limit problem gambling. On the other hand, problem gamblers are likely to be more willing to wait in a queue and stay longer when they do.

Figure 11.4 The impact of lifting caps in NSW hotels



This is not to argue that constraining gaming machine numbers is a completely ineffective strategy for reducing problem gambling. Rather, its efficacy will depend on a number of factors, including the number of machines involved, the size and proximity of venues, offsetting increases in machine utilisation and the intensity of play, and the behavioural responses of problem gamblers. The Productivity Commission’s analysis suggests that venue caps have some advantages over jurisdiction-wide caps. But more targeted mechanisms for harm minimisation would involve less collateral disadvantage to recreational gamblers and would be more effective in reducing social costs. Although caps have popular appeal and apparent administrative simplicity, they are, at best, a blunt instrument for dealing with the social problems posed by gaming machines.

11.5 Consumer protection

In canvassing more cost effective (or ‘market friendly’) consumer measures, the Productivity Commission focused on those that would *widen*, rather than narrow choice. It identified three key areas where there was untapped potential for reducing social costs without significantly diminishing the benefits available to recreational gamblers:

- informed consent;
- controls *for* consumers; and
- restrictions on games and/or venues (figure 11.5).

Figure 11.5 Consumer protection measures

Informed consent	<ul style="list-style-type: none"> •Price information •Expenditure statements •Warnings •How games work •Ethical promotion
Controls for consumers	<ul style="list-style-type: none"> •Pre-commitment (eg <i>user</i> imposed spending limits, self-exclusion and time limits) •Filter software on PCs (limiting access to unsafe internet sites) •Coordinated venues/sites (so that pre-commitment is effective)
Venue and game restrictions	<ul style="list-style-type: none"> •ATMs and credit restrictions •Limits on bill acceptors •Limits on machine intensity •Enforced breaks in play •Limits on the form of payouts

The first two are wholly about expanding the basis for choice and, thus, enhancing ‘consumer sovereignty’ — both for recreational and problem gamblers — whereas measures in the third category involve some tradeoff. For this third category, attention was given to aspects of choice that exacerbated the risks of problem gambling, but were not generally important for recreational gamblers. However, most would require pilot testing and monitoring to determine their cost effectiveness and most appropriate design.

Informed consent

The principle of ‘informed consent’ should apply with particular force to the gambling industries, given the potential for consumer losses. But the Productivity Commission found a lack of basic information about the price and nature of some gambling products, let alone the potential dangers from ‘excessive consumption’.

Meaningful ‘price’ information

Unlike many other consumption items, the price of gambling is often hidden. The cost of purchasing a lottery ticket or the money put into a gaming machine is not the price of gambling on these forms. The price is the money outlaid, *less* the expected value (statistically) of the prize. But few people know the odds of winning, so that the price is unknown. For example, few poker machine players would know that the chance of getting say ‘five rhinos’ in *Black Rhino* would be only one in ten million — even less than winning the lottery.

Nor do minimum payout ratios for gaming machines convey much information about likely spending rates. For example, the *Fast Fortune* machine comes with alternative rates of return of 87.7, 90.04, 92.33 and 94.99 per cent — all above the statutory minimum. These different rates are not disclosed to consumers and cannot be inferred reliably, except by careful observation and playing tens of thousands of games. The implication is that the price on these otherwise identical machines could vary by as much as 146 per cent, without the consumer knowing.

While there are complexities involved, the Productivity Commission saw considerable scope to provide more meaningful information about the effective ‘price’ of playing poker machines and lotteries, including the likelihood of receiving high paying winning combinations. The prime objective of better information is simply to empower consumers, not to deal with problem gambling. However, better information about the odds and average costs of gambling might help reduce the false perceptions that can underlie problem gambling.

Statements of expenditure

Many gamblers are not only ignorant of the true price of their preferred games, but also appear to have poor awareness (or biased recollections) of their losses relative to their wins. This has been borne out by surveys. (For example, the ABS *Household Expenditure Survey* indicates gambling expenditure in 1993-94 was about one-quarter of the actual level.) As a result, a potential early-warning signal about problem gambling is muted.

One advantage of the internet as a gambling medium is that it provides a gambler with ready access to comprehensive information about his or her spending pattern and levels. There is scope to provide more such information within other gambling forms. This has been facilitated by the advanced information technology in gaming machines. Accounts could already be made regularly available to existing holders of ‘loyalty cards’ in casinos and clubs, and to TAB account holders. If the industry can collect information on consumers’ spending patterns, it seems only reasonable that information also be divulged to consumers themselves.

How games work

Erroneous beliefs about what determines ‘success’ in gambling are legion. Apart from leading to poor decision-making by the average consumer, these false perceptions are seen by psychologists as major contributors to problem gambling.

One of the more widespread misconceptions (evident in problem gamblers’ frustrated ‘chasing of losses’) is the notion that gaming machine payouts depend on

previous outcomes from a machine. Information in this area is essential to inform consumer choice and again could help prevent the development of gambling habits and attitudes that lead to problem gambling. Such information could be made readily available through pamphlets and signs (in a range of languages) — as is done now in casinos to introduce people to the rules of table games. Indeed, the Australian Gaming Machine Manufacturer’s Association (AGMMA), the organisation representing gaming machine manufacturers, has recently produced a booklet to distribute to venues about common fallacies.

‘Health (or wealth) warnings’

As noted, a variety of people within society can become problem gamblers. The incidence and continuum of impacts, and the costs that each problem gambler can impose on others, define it as a public health issue. That in turn implies the need for better and more readily available information for consumers about:

- the risks and consequences of excessive gambling;
- the signs of an emerging problem; and
- sources of assistance and advice.

Such information needs to be disseminated within the general community, as well as at gambling venues. It needs to be much more accessible than is currently the case in many venues. The effectiveness of many of the warning signs currently in use is doubtful (eg ‘Have fun, but play it safe’). They are rarely as explicit as in other areas of public health (eg ‘Speed kills’).

Advertising and promotion

As with other products, gambling advertising generally accentuates the positives and ignores the negatives, so as to stimulate consumer demand. In this case, that can be hazardous. As one former industry leader commented at the Productivity Commission’s public hearings, ‘all advertising for gaming is misleading because it only shows people winning. That is not the experience of most people in gaming’. The message is that everyone can be a winner. This can help foster the sorts of erroneous beliefs about gambling that lead to problems. As noted, unlike most other products, it is difficult for many consumers to learn the truth for themselves.

The voluntary codes examined by the Commission are useful, but none require the ‘warnings’ that apply to other public health issues, or target the misconceptions.

Controls for consumers

Since many gambling problems stem from impaired player control and lack of awareness of accumulating losses, the question arises as to whether gaming technology could provide consumers with the ability to self-constrain their future actions, if they *wished* to do so. This is the idea of ‘pre-commitment’.

Currently, apart from internet gaming sites, pre-commitment options are limited to relatively cumbersome, embarrassing and (circumventable) arrangements whereby players can ‘self-exclude’ themselves from venues. Such arrangements appear to have operated most effectively in casinos, where there are statutory provisions protecting the venue from liability and enabling relatively simple contractual arrangements. There is a case for extending statutory cover to all venues, while making it mandatory to advertise the facility and for venue management to act on all requests for self-exclusion.

But technological advances open up the possibility of going well beyond such ‘all or nothing’ choices, to include self-imposed limits on gambling duration and expenditure. These pre-commitment features are already available on Australia’s first regulated casino gaming internet site (www.lasseters.com.au). Indeed, these and other information-rich features influenced the Productivity Commission’s judgement that ‘managed liberalisation’, rather than attempts at prohibition, was the most reasonable policy option for internet gambling.

In the case of ‘real’ gambling, the advent of smart card technology provides scope for equivalent protection. Such technology is already being contemplated in a number of jurisdictions. It is not clear, however, that the industry sees smart cards as a *two-way* mechanism for providing information and control. It is critical that pre-commitment options be an essential feature of such technology. The scope for achieving coordination among venues also needs to be given serious consideration.

Modifying venues, game features and design

The disclosure and pre-commitment strategies just described have considerable potential to reduce risks of problem gambling, while at the same time enhancing the basis for effective consumer choice. Changes to venues or game design could also help, but would reduce consumer choice somewhat. Their net benefit would depend on the extent to which they can reduce the costs of problem gambling relative to any diminution in the entertainment value for recreational gamblers.

Access to cash and credit

It is generally illegal for venues to provide credit to gamblers and many do not cash cheques — on the basis that such facilities pose major risks to those who have difficulty controlling their gambling. By similar logic, casinos have taken a lead by situating their Automated Telling Machines (ATMs) away from the gambling floor. Problem gamblers surveyed by the Productivity Commission ranked ATM location as one of the most important issues for effective harm minimisation.

The Commission's *National Gambling Survey* found that problem gamblers were much more likely than others to withdraw money from an ATM while playing the pokies (table 11.2). Thus, the collateral effects of constraints imposed on access to ATMs are likely to be small. Further evidence of this is the fact that banks generally do not place ATMs in hotels or clubs without poker machines.

Table 11.2 How often do you withdraw money from an ATM at a venue when you play the poker machines?

	<i>Never or rarely</i>	<i>Often or always</i>
	%	%
Recreational players	90.0	4.6
Problem gamblers (SOGS 5+)	47.0	37.8
Problem players (SOGS 10+)	25.2	58.7

SOGS (South Oaks Gambling Screen): a score based on the answers to a set of questions used to determine whether a person is a problem gambler (PC 1999).

Source: PC *National Gambling Survey*.

An outright ban on ATMs in venues with gambling is one option, although it would maximise any inconvenience to recreational gamblers and other patrons. More targeted and potentially cost-effective options include restrictions on the location of ATMs and lower withdrawal limits in gambling venues.

Gaming machine modifications

As noted, gaming machines have been the major source of problem gambling. In addition to their wider availability, contributing factors include their continuous nature, the ability to increase the size of successive bets and the structure of payouts. Unlike some other forms of gambling, such as racing or casino games, the fact that these games involve computer-style machines increases the technological possibilities for reducing hazards through design modifications.

The Productivity Commission canvassed a variety of options in its report. Many of these could be programmed into the machines to allow interaction with the gambler. They include:

- curtailing or eliminating the use of ‘bill acceptors’ on machines;
- limitations on spending rates;
- enforced breaks in play at convenient intervals (such as after a significant win); and
- big payouts being made only by cheque (as currently occurs on leaving a casino).

While such measures appear likely to have a beneficial effect for problem gamblers, their impacts on recreational gamblers would need to be assessed.

Less promising options

A number of other measures proposed by participants were found to lack sufficient prima facie evidence to warrant immediate consideration, or were simply judged unlikely to be of net benefit. These include:

- altering the lighting and sound effects for machines;
- increasing the minimum duration between button presses;
- requiring clocks in venues;
- banning alcohol consumption in gaming areas;
- eliminating all linked jackpots; and
- having a statutory requirement to identify (and exclude) problem gamblers.

Nor is it the case that just *any* realisation of the harm minimisation mechanisms with some prospect of efficacy would be appropriate. For example, requiring a cheque to be issued for winnings over \$50 would be burdensome for many recreational players, while it may be quite cost-effective for a threshold of \$250. Similarly, enforced breaks that occur too frequently would be likely to involve net costs and, in the absence of smart card technology, be circumvented by simply switching machines.

Self-regulation versus statutory regulation

With the escalation of public concern and threats of government action, sections of the gambling industries have developed codes of practice covering such areas as advertising, signs about risks and counselling services, the training of staff and

responsible serving of alcohol to gambling patrons. Most of these are recent (or yet to be implemented) and their efficacy is largely untested. However, industry representatives argued that such an approach would be the most effective, as venues in a competitive industry have a strong incentive to look after the best interests of their clients.

In principle, self-regulation has important potential advantages over much ‘black letter’ regulation in that it can allow for flexibility (and learning) and typically specifies objectives and options rather than prescribing the means for achieving the objectives. That way, each venue could, in theory, determine the minimum cost approach for achieving the regulatory objective.

However, the self-regulation approach works best when venues have strong incentives to meet clearly defined objectives, and outcomes can be easily monitored (eg train safety and effluent discharge levels). These conditions do not apply to problem gambling. Enforcement of either common law or statutory duty of care could prove more difficult than in other situations — because of the scope to shift the ‘blame’ to other venues or gambling modes.

Importantly, the incentive to comply is also likely to be compromised by the substantial proportion of takings derived from problem gamblers (averaging an estimated 42 per cent of revenue from poker machines). Possibly reflecting this, compliance with existing voluntary codes has clearly been inadequate in some cases. (For example, printed notices about risks and sources of advice on problem gambling are often hard to find, whereas ATMs notionally located ‘away from the gambling floor’ are often all too easy for gamblers to find!)

On these grounds, and notwithstanding the in-principle attractions of light-handed regulation, the Productivity Commission saw a case for prescriptive regulations in some key areas, rather than simply relying on voluntary codes of practice. Since the Commission completed its draft report, regulatory initiatives of this kind have been signalled in a number of jurisdictions, including the New South Wales *Gambling Legislation Amendment (Responsible Gambling) Act 1999*.

11.6 Towards improved regulatory processes

Many of the deficiencies and gaps in current regulations are the legacy of the way policies have been made. While some jurisdictions have done better than others — and all jurisdictions are currently re-appraising aspects of their regulatory processes and institutions — there have been widespread deficiencies. These include:

-
- lack of rigour in assessing the costs and benefits of alternative options, including muted incentives to make such assessments;
 - confusion in the differing roles of institutions engaged in policymaking and advice, reflecting the separation of gambling modes and the relative importance given to revenue raising;
 - lack of community consultation about attitudes to, and the possible social and other consequences of, regulatory decisions;
 - lack of transparency in decision-making and in the provision of information about impacts; and
 - little systematic monitoring and evaluation of the consequences of such decisions, once implemented.

Good regulatory process requires that decision-makers have the appropriate degree of independence and control, that their objectives are clear and their decisions well-informed, and that the basis for their decisions is transparent and publicly accessible. Such features are especially important in a policy area such as gambling, which is characterised by conflicting pressures and incentives for government — and the potential for major winners and losers, within business and the community, from different regulatory outcomes.

There are several distinct functions that need to be performed in any proper regulatory environment for the gambling industries, of which three key ones are policy development, ‘control’, and enforcement.

Policy development

Some policy decisions are properly the preserve of parliaments, with ministerial responsibility for their development. These include threshold decisions bearing on the extent and accessibility of different gambling forms — including caps, allocation rules and ownership decisions. Some important ingredients of good process at this level include:

- a requirement for regulation impact statements and the procedural steps that underpin them, with independent public reviews to identify costs and benefits of different options; and
- public consultation about options under consideration.

Independent ‘Control Commissions’ in each jurisdiction

The resulting legislative standards within each jurisdiction need to be well administered. This involves such important decisions as who gets licences to offer different gambling activities, who may work in different venues, appropriate technical standards and when penalties apply for breaches of licence conditions and regulations (including any consumer protection and harm minimisation requirements).

Good process at this level demands genuine independence of the regulator, which must exercise any discretion in an impartial manner, without undue influence or interference by industry, community or the government of the day. The core regulatory body, therefore, should have a statutory basis, with authority vested in tenured commissioners who have no connection to the industry and report to Parliament through the relevant Minister(s) in each jurisdiction.

The core regulatory body should also have:

- jurisdiction over the gambling industries as a whole (to allow more coherent decision-making in relation to the different modes);
- an advisory role to the Minister and Parliament on major public interest matters, including those requiring legislative change;
- a requirement to consult with local communities in making decisions about licence applications; and
- statutory guidelines which make explicit the over-riding objectives of consumer protection and the public interest.

Regulatory enforcement

The enforcement function is about ensuring that venue operators comply with licensing conditions and other regulatory requirements. Under regulatory provisions for harm minimisation, compliance with these requirements would be part of this function.

This ‘policeman’ role is generally regarded as needing to be undertaken separately from the control function (the ‘judge’) — a principle embodied in the so-called New Jersey model of gambling regulation. Independent Pricing and Regulatory Tribunal (IPART) endorsed the principle in its recent review in New South Wales. The Productivity Commission also saw benefits in this separation, including from public confidence in the implied checks and balances.

11.7 Conclusion

The Productivity Commission's inquiry found some fundamental flaws in the regulatory regimes for the gambling industries. Since then, a number of governments, most notably New South Wales, Queensland and Victoria, have introduced initiatives to bring greater balance to the regulatory environment. These have revealed an acceptance of the need to address consumer protection and to achieve better informed regulatory decisions. Some examples include the removal of requirements for the regulator to foster tourism and economic development, the introduction of a requirement to consult with communities, constraints on deceptive advertising, limits on bill acceptors and the possibility of using smart cards with pre-commitment options for gaming machines.

While there are *prima facie* grounds for many of the new consumer protection measures, it is unclear to what extent there has been provision for trialing and independent evaluation to ensure their effectiveness. Most governments have also done little to deal with the anticompetitive features of their regulatory regimes — partly because, by their nature, these features tend to be locked in through contracts. There have also been limited changes to governance arrangements for regulatory administration.

Getting the regulatory framework right is the key not only to better social outcomes, but also to the gambling industries' own future. Under an appropriate regulatory framework the industry is still likely to prosper, although not necessarily with the rates of growth and profitability attained in the 1990s. But any such relative decline needs to be considered against some plausible 'counterfactuals'.

This paper began by alluding to the death of a baby. In the United States, a similar incident in August 1999 shocked South Carolina. South Carolina is now moving to prohibit all gaming machines by mid-2000 — testimony not just to a single incident, but also to a long period of mismanaged regulatory policy in that State. That regulatory response may eliminate many of the social harms of gambling, but it will do so at the cost of also eliminating the consumer benefits.

Bans and quotas are a signature of failure to address, in a balanced way, the complexities of gambling regulation. The challenge for Australian regulators is to find targeted means of preventing or ameliorating the harms from problem gambling while preserving the consumer benefits that these industries can offer. The Productivity Commission's 'market friendly' menu was designed to assist them in finding that balance.

References

- ACIL Consulting 1999, Submission to the Productivity Commission Inquiry into Australia's Gambling Industries, Submission no. 155.
- O'Donoghue, T. and Rabin, M. 1999, 'Doing it now or later', *American Economic Review*, vol. 89, no. 1, pp. 103–24.
- ORR (Office of Regulation Review) 1998, *A Guide to Regulation*, 2nd edn, AusInfo, Canberra, December.
- Pandazopoulos, John, the Hon. 2000, *Responsible Gaming: A Consultation Paper*, Minister for Gaming, Victorian Government, March.
- PC (Productivity Commission) 1999, *Australia's Gambling Industries*, Report no. 10, AusInfo, Canberra.
- Tattersalls 1999, Submission to the Productivity Commission Inquiry into Australia's Gambling Industries, Submission no 156.

Discussant — *David Johnson*

Melbourne Institute of Applied Economic and Social
Research

One of the positive outcomes of this work by the Productivity Commission (PC) is the provision of information. The information provided in the draft and final reports offers an enormous amount of material for people wishing to understand the gambling industry. Indeed, the PC has been very successful in highlighting several issues with important implications. Rather than dissecting the PC's report or Gary Banks' presentation, this rejoinder highlights some of these important implications.

D11.1 A market friendly approach to regulation

In its draft and final reports on gambling, the PC claims to utilise a 'market friendly' approach to the social problems associated with gambling. The premise behind the 'market friendly' approach is that it is the government's role to overcome market imperfections and failures. Society's overall objective of maximising social welfare will be best achieved in a situation where goods and services are transacted in as close to a market-like environment as possible. This view sees the efficiency of the unfettered market delivering the highest national product. Government then redistributes this income to maximise social welfare.

According to the New Palgrave Dictionary of Economics, 'regulation' consists of government actions to control price, sale and production decisions by firms in an attempt to prevent private decision-making that would take inadequate account of the 'public interest'. It is the interpretation of 'public interest' that is at the heart of Gary's suggestions for gambling regulation.

The rationales for government involvement with the gambling industry are fourfold:

- to remedy perceived failure of the market to operate in a competitive way — this might occur because of information problems or externalities which impose costs on parties other than those involved in the transaction;
- to encourage growth and development in particular industries;
- to raise revenue; and
- to protect vulnerable consumers.

All of these rationales have been relevant to government involvement in the gambling industries at some stage. However, in recent times the latter rationale has been most important.

Social impacts

The PC has attempted to define the problems of gambling in its inquiry report. These seem to revolve heavily around the idea of social impacts.

The social impacts of gambling have come to mean the negative outcomes associated with over-indulgence in gambling, the negative association of gambling with criminal activity, and a general and pervasive moral laxity emanating from the promotion of what is seen to be a socially harmful activity. However, the PC points out that these effects must be balanced against the revealed preferences of a very large proportion of the adult population who gamble in a controlled way and derive considerable utility from this consumption.

Concern about social impacts is also partly a reflection of the times. The current 34 quarters of growth has been accompanied by a widening disparity in the distribution of income within Australia and, thus, a greater role for government to ameliorate the negative effects of this gap.

Finally, social concerns are a function of the nature of the product and its consumption. Some of the pertinent facts about gambling are as follows.

- Gambling is widely accepted in Australia, with about 80 per cent of the population purchasing it. For most people there is a clear preference for gambling over alternative entertainment products.
- The price of the commodity is not necessarily clear to the purchaser.
- It is often regarded as a demerit good, because some consumption can be addictive and over-use can have very bad consequences for these persons.
- Gambling is largely a State responsibility and, therefore, with 6 States and 2 Territories there are 8 different jurisdictions regulating gambling in Australia.
- Gambling is heavily taxed, the tax is regressive and the tax is an important revenue earner for government.
- The gambling industry is vulnerable to criminal involvement and, therefore, requires processes to ensure probity and to minimise conditions in which criminality might flourish.

The PC Recommendations

Gary Banks and the PC make a strong case for government involvement in the industry. Their recommendations are not heavy-handed and entail a smaller, rather than a larger, involvement. They accept that there is a strong consumer demand for gambling and there is no reason why this should not be satisfied. However, they

also accept that providing protection for the vulnerable is an important role of government and this is the focus of their recommendations.

Gary Banks questions the revenue-raising role of government in relation to gambling. He implies that the importance of gambling as a revenue raiser unnecessarily compromises government and, in some cases, has led to inappropriate or poorly developed policy.

Gary suggests a regulatory framework that aims to maintain the social benefits, but to minimise the social costs. However, since it is judgement that determines the balance of benefits and costs in assessing the net effect of gambling, the policy remedies to minimise social costs also involve a considerable amount of judgement.

The general thrust of the recommendations is to foster an environment where vulnerable gamblers are more aware of the dangers of their activity and are empowered to limit their consumption. The recommendations maintain the ‘market friendly’ philosophy, but these regulations also differ in a fundamental way from most other regulation of the industry.

D11.2 Consumption focus of gambling regulation

Consider the nature of the gambling regulations proposed by the PC. Some of the proposed regulations focus on the actions of the producers. Proposals cover the removal of unwarranted exclusivity clauses for some producers in some jurisdictions and maintaining or improving probity measures for producers. Most of the proposals, however, are directed specifically at consumers and the consumption process. This is in sharp contrast to most regulatory frameworks, which are directed more at production and the nature and structure of producers.

It is certainly possible to define some of the problems of gambling in terms of market failures. For instance, the costs of problem gamblers on their friends, relatives and on community services (but not on themselves) can be interpreted as negative externalities. The general social dysfunction and criminal activity attributed to, or associated with, gambling can be visualised as a negative externality.

Similarly, information problems may be detected — excessive consumption may be regarded as a result of a lack of information or misinformation (eg advertising that everyone’s a winner).

However, the rationale for the regulatory framework doesn’t really get to the heart of the issue. The key issue, as will be evident below, is that the major part of the social costs emanate from a small group of vulnerable consumers, known as

problem gamblers. Problem gamblers, variously estimated to comprise up to 2 per cent of the gambling population, incur about a third of total gambling expenditures (losses). In the process, individuals, their friends and relatives endure great hardship and use community resources. Even if it were possible to solve all of the market failure problems, society would not eliminate problem gambling and, hence, the perceived social dysfunction associated with it. Thus, the more appropriate way to view the regulatory framework for gambling is outside of the conventional production-focused situation.

The PC's quantification of social benefits and costs can be used explicitly to define the objectives of regulation.

Social benefits and costs

In reviewing the net effect of gambling, the PC attempted to quantify and compare the social benefits and costs for consumers and, hence, society, since most of the production-related effects (such as effects on growth and employment) are small. Given the high degree of subjectivity involved in such an exercise, the sign and size of the resulting net benefits from gambling cannot be completely assured, owing to the large margin of errors involved. While the benefits can be reasonably estimated from consumer surplus calculations using estimates of consumer elasticities, the costs depended crucially on subjective views about the level of harm entailed in problem gambling.

In the calculation of the net benefit, the PC implicitly summed across individuals. The costs, which are largely incident on one class of consumer (specifically problem gamblers and their relatives and friends), are deducted from the benefits which accrue to a different class of consumer — gamblers in general.

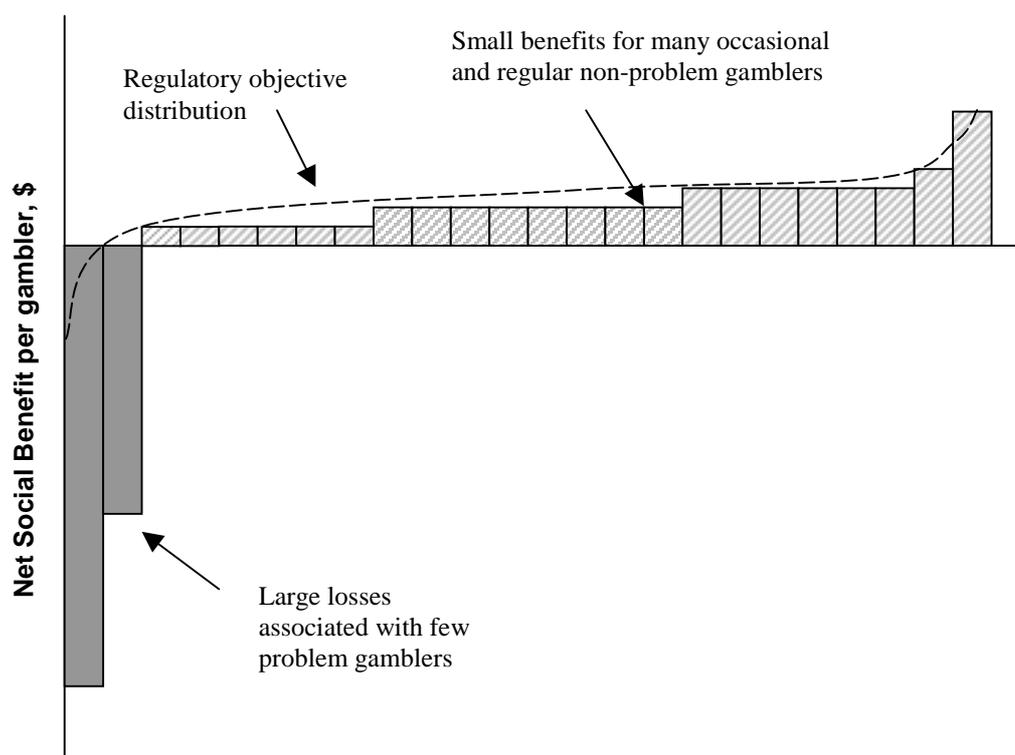
The PC calculated the positive effects from various measures of consumer surplus based on elasticities estimated for various forms of gambling. While the precise elasticities are unknown, the PC uses a range of plausible elasticities to generate a range of estimates of consumer benefit.

The negative effects are estimated in some detail from a range of sources. However, they are mainly attributable to the large losses of problem gamblers. They included costs associated with divorce, family break up, depression, suicide and lost productivity. The social costs focus on a relatively small number of people, while the population at large enjoys the benefits of gambling.

The main reason for regulation is to reduce the losses of the small number of problem gamblers, which will eliminate a large part of the negative social effects of gambling.

The following figure depicts the regulatory objective of changing the consumption profile. The adult population of gamblers in Australia has been ranked individually from those who derive the most benefit from gambling (according to social criteria) to those who incur large social costs.

Figure 11.6 The distribution of social benefits from gambling



The result is a very skewed bar chart with 1–2 per cent of gamblers incurring very large losses, while most of the population enjoys modest welfare gains. Note that for most people, the net average losses over a long period are reasonable measures of their welfare gain. The intuition is that those who are not problem gamblers know that over time they can expect to make a net loss, but that the enjoyment of the environment, its ambience, the associated social activity, the thrill of winning and so on, more than compensate for their net losses.

The objective of regulation is to produce a consumption profile more like the dashed line. Here the small gains for many are maintained, while eliminating the large losses for problem gamblers.

D11.3 Some consequences of a consumer focused rationale for regulation

This rationale raises some interesting consequences for gaming regulation, including:

- government revenue;
- consumer sovereignty;
- implications for producers; and
- the measurement of success.

Government revenue

If the main aim of government is to change the consumption profile, the government may use its taxing powers to achieve this task. However, the efficacy of taxation as a tool to change consumption pattern depends on the elasticities — not of consumers in general, but of the actual and potential problem gamblers. The evidence regarding elasticities seems inconclusive, although the PC seems to believe that they are likely to be low, rather than high, for consumers in general. They are likely to be quite low for problem gamblers. If this is the case, high taxes are not effective in deterring gambling and simply act to exacerbate the bad effects of problem gambling.

There is also the issue of distribution. Gambling taxes are highest for lower income groups, who spend more of their discretionary dollar on highly taxed gambling products.

Consumer sovereignty

The stated rationale for changing the consumption pattern attacks consumer sovereignty and reduces individual freedom. This contradicts recent trends in social policy, which have seen a concerted program of liberalisation with regard to other seemingly harmful, self-administered social activity. For instance, Australia has liberalised laws concerning alcohol usage, increasing acceptance of illegal drugs and increasing acceptance of victimless illegal sexual activity. Interestingly, liberalisation of social policy has occurred concurrently with liberalisation of economic policy.

In view of these trends, it seems unlikely that there would be public support for strong measures, such as prohibition or even more stringent quantitative restrictions on activity (such as caps on numbers of poker machines). Nevertheless, the public has clearly viewed, with great concern, the social problems arising from gambling abuse. What seems to be required is a measured approach with some restrictions.

Implications for producers

To achieve the stated objectives of regulation, a change in the profile of consumption is necessary. This potentially has more profound implications for producers than regulation aimed at freeing markets. Problem gamblers provide a third of the income of producers and 42 per cent of the income of gaming machine operators. Unless other consumers can be persuaded to gamble more, even modest success of the regulations would imply a large contraction in the size of the industry.

Measurement of success

Measuring the success of regulations, as outlined by Gary Banks, is likely to be difficult. As suggested, there are potentially large errors in the calculations of the net benefit shown in the PC report and in the graph.

The success of policies can be measured with respect to their own aims, and also more generally with respect to societal norms. The first measure is an internal consistency check and may consist of plotting the dashed line against the bar chart. If at some period in the future, after new regulations have been introduced, the consumption profile changed to the dashed line then the regulation might be judged successful.

However, there is a second and more telling criteria. Here, the success is not measured against the initial objective set by legislators, but against societal norms. In this evaluation, the judgement of the legislator and the extent to which they can sense, or measure, existing societal norms is critical. However, even if a judge is able to accurately tap current opinion, there is no guarantee that societal norms will not change. Indeed, the very success of a regulatory program may lead to changes in the expectations of society.

General discussion

The discussion focused on:

- whether there were external social costs associated with gambling; and
- several issues relating to internet gambling.

Social costs

One participant queried the claim that not all the costs of problem gambling were internalised. Most people had some compulsive tendencies, be they workaholics, over-eaters, compulsive thrill seekers, drinkers or entrepreneurs. Typically, we relied on individual and family responsibility to check these tendencies.

Gary Banks responded that the key issue was the addictive dimension of gambling and whether people were making decisions in a rational way. The notion of ‘policy-relevant’ costs was perhaps more useful in this respect than just external costs. What also distinguished gambling from some other compulsive activities was the extreme nature of some of the costs — for example, compulsive cinemagoers were unlikely to commit suicide. The costs of emotional trauma, depression and conflict associated with problem gambling had to be offset against the benefits of gambling, but these costs did not show up in market prices. There was also little evidence of these costs being factored into implicit family contracts. Instead, problem gamblers typically concealed or lied about their behaviour.

Internet gambling

Several comments were made with respect to internet gambling, control mechanisms and likely government responses.

It was noted that, often in areas where there was a morality dimension or a perceived social cost, the initial response was a call for a ban rather than more appropriate regulation. Indeed, this had been the case with many of the problems associated with the internet. Gary Banks responded that the report showed that a ban on internet gambling was not a sensible policy approach, because it would violate the over-arching principle of targeting the costs of gambling while not curbing the benefits.

The issue of how to control foreign providers of internet gambling services was discussed. The Productivity Commission report argued that by allowing licensed internet gambling sites, consumers would be more likely to comply with bans on unlicensed sites, and this paralleled the way ‘real’ gambling was regulated. It was

easier to screen for illegal gambling sites than to screen for illegal content (eg pornography). Current control mechanisms (eg filtering IPP addresses) would not support a complete ban on internet gambling, but would support control over unlicensed sites. But control mechanisms became less critical if there were legal, licensed sites available to consumers. And since consumers would be unlikely to use sites where there was a high probability of not collecting their winnings, good sites could drive out bad.

12 Regulating culture: has it ‘Gone with the Wind’?

Patrick Messerlin

Institut d’Etudes Politiques de Paris

12.1 Introduction

Audiovisual services cover a wide range of services, ranging from art to information. They include:

- motion picture and video tape production and distribution services;
- motion picture projection services;
- radio and television services;
- radio and television transmission services;
- sound recording; and
- other audiovisual services.¹

This classification covers ‘content’, not the physical infrastructure required for transmitting the content (generally included in telecommunications). Most of these services are closely associated with political power, both on a daily basis (radio or television interviews and news) and in their ability to create collective symbols. This explains why audiovisual services generated such controversy during the Uruguay Round, and why negotiators are so cautious about introducing them into the agenda for the new round of World Trade Organization (WTO) talks.

The key issue raised by audiovisual services is their ‘nature’. Are audiovisual services ‘industrial’ activities, like all other services, or are their links with culture deep enough to protect them behind a cultural ‘exception’ or ‘specificity’?

¹ Both the UN *Provisional Central Product Classification* and the GATS *Services Sectoral Classification List* place audiovisual services in the *Communication services* group, along with its four other sub-groups — *postal, courier, telecom, and other services*.

Economists are well versed in the best disciplines to impose on industries, and there is no reason to exempt ‘industrial audiovisuals’ from these disciplines. Some of these disciplines are trade-related, such as opening the domestic market to foreign competition (cross-border and establishment) in a non-discriminatory way, including affording national treatment. Other disciplines to be imposed on industries involve domestic regulatory reforms, including:

- the creation of appropriate markets for intellectual property rights;
- the regulation (or otherwise) of interactions between various services (for instance, should there be constraints on time ‘windows’ for showing films on screen, on television, on videocassette, etc, or should audiovisual producers be free to decide?);
- enabling network access (airwaves, cable or digital television); and
- enforcing competition laws.

By contrast, handling ‘cultural’ goods or services is a more difficult task. First, it requires the scope of cultural goods to be defined. This is not a simple process, as it may face strong prejudices and is often subject to powerful vested interests able to capture ‘culture’ for their own profit. Second, if existing disciplines on the use of certain instruments, such as subsidies, are relaxed for ‘cultural’ goods or services, this may require more subtle ways of reconciling the freedom to create and the freedom to trade. In particular, it would place a premium on more efficient domestic regulation, that is, regulation that supported culture successfully, while requiring less direct public intervention.

The main message of the paper is that ongoing economic and social changes are increasingly separating the industrial and cultural segments of audiovisual services. This increasing separation offers the opportunity to combine a better policy for industrial audiovisuals (based on progressive liberalisation) with better ‘regulation’ of cultural audiovisuals (institutions and instruments for supporting culture at its source, rather than through the platforms by which culture is disseminated).

Section 12.2 examines the relationship between the industrial and cultural aspects of books and paintings — two main platforms for the dissemination of culture that have been subject to open trade and a free market policy for the last half-millennium. Section 12.3 describes the protection of audiovisual services in the European Community (EC). Section 12.4 summarises its consequences. The most important one is that protection has been (and still is) *much more* harmful to the cultural than to the industrial segment of the audiovisual sector. Section 12.5 analyses the limitations of the current WTO legal framework (GATT and GATS) for addressing these issues. Hence, it suggests the need for a Reference Paper on audiovisual services to interpret WTO disciplines in a way that encompasses the

most important aspects. Section 12.6 concludes by making a few remarks on the current environment for future WTO negotiations in audiovisual services, on the content of the proposed Reference Paper and possible domestic regulatory reforms.²

12.2 The nature of audiovisual services: To be cultural, or not to be?

Economists often ignore the nature of the relationship between audiovisual services and culture, as they feel uneasy dealing with the issue. Such neglect is a recipe for policy failure. In past and current debates in the WTO and other international forums, vested interests have sought to tie audiovisual services to culture as tightly as possible, in order to retain the status quo. The last decade shows that this has been a successful tactic.

In the coming decade, culture-minded people might join economists in looking at the issue, given the consequences of the existing protectionist EC policies (see section 12.4). This is best illustrated in France, one of the EC Member states having most tightly linked audiovisual services to culture. Since the summer of 1999, mounting discomfort has seen the French press pinpointing flaws in French film policy in an unusually aggressive tone, and highlighting the ‘privilèges’ (rents) drawn from this policy by a few vested interests — privilèges felt all the more unwarranted because the policy shows obvious signs of failure (Messerlin and Cocq 2000). French politicians have yet to reflect this discomfort. This is not surprising. The stakes are so high — political careers in a media-intensive society depend on audiovisual services — that politicians are unlikely to enter the debate before the outcome is clear.

Culture, books and paintings

During the last half-millennium, books and paintings have been major platforms by which culture has been disseminated. They offer interesting lessons for audiovisuals about the relationship between the industrial and cultural segments.³

² As these topics are tightly related to the political life of a country, this paper focuses on France — the author’s home country. However, many of the observations about France could easily be extended to other countries that follow protectionist policies in the name of culture.

³ Architecture and large sculptures are left aside because they are almost non-tradable. There are, however, some notable exceptions — a few museums have reconstituted Pergamon and other places, and art ‘razzias’ (raids) by war victors have occurred since the beginning of human history. If books are tightly related to language, paintings are not. Audiovisual services lie somewhere in between.

From the fifteenth century until the 1920s, the fact that books (and the press) were the dominant platforms for the dissemination of culture did not prevent a very liberal trading regime. Imports and exports of books were free from quotas (at least, in democratic regimes), and subject only to the few exceptions valid for all other goods under GATT Article XX, which deals with public morals and order. Imports of books were subject to non-discriminatory and relatively moderate tariffs. Imported books enjoyed national treatment. That is, they were subject to the same tax regime as domestically produced books and to the same pricing regime (many countries still impose a minimum retail-pricing regime, despite its pro-collusive effects on domestic publishers). Lastly, publishers have rarely been subsidised since printing became an industrial activity. Only very few books — collectors' items — have been systematically subject to trade constraints (quotas or bans), but not before the late nineteenth century in Europe.

During this liberal regime, the European industrial book sectors prospered and European cultures flourished as never before. Publishers based in large countries thrived, as did those working mostly for export (because their national language did not allow a vast readership, publishers such as the Dutch shifted successfully from Latin to French and then to English). No European country lost its capacity to produce industrial and cultural books, and many without such a capacity five centuries ago developed it.

The same liberal trade regime also applied to paintings. Except for a few masterpieces, constraints on trade in paintings were almost non-existent until the late nineteenth century. In fact, the largest collectors only began to have systematic lists of their art works during the nineteenth century (those interested in art history, such as Vasari, were exceptions). In European countries, severe trade constraints on masterpieces were imposed after the 1870s, as European private and public collectors were less inclined to spend as much as the more enthusiastic new collectors from outside Europe (from the United States since the late 1800s and from Asia since the late 1900s). And, in its first (1968) *Italian Art Treasures* ruling, the European Court of Justice has reminded us that, despite their uniqueness, works of art qualify as 'goods' within the meaning of the Treaty of Rome because they are products having a monetary value, and as such, may be the object of commercial transactions.

These two brief surveys suggest two lessons.

First, they do not support the *general* argument that the industrial segment of an activity should be protected in order to ensure a lively cultural segment. Support, if any, for this argument needs to rely on some *specific* economic features of the audiovisual sector. From this perspective, painting deserves special mention. It has not benefited from an industrial evolution comparable to that for books with

printing. Painting technology has evolved in terms of colours and materials.⁴ But this evolution has not been based on industrial changes such as cost-cutting mass production, and it has not affected large paintings. As a result, large paintings have attracted more subsidies (eg public orders from the Church, princes and democratic governments) and have become disconnected from market rules. Nevertheless, there are two additional lessons from the history of painting that are useful in designing the best domestic regulatory frameworks to support culture in audiovisual services. The first is the importance of economically sound relationships between painters on one hand and patrons, sponsors or galleries on the other. The second is the perverse impact of the managed trade in masterpieces, with fantastic rents associated with speculation in masterpieces subject to the restrictions.⁵

Another broad lesson from both the book and painting sectors is that they have been divided into two clear-cut segments:

- a large segment of fully industrial activities operating under free market and trade disciplines and able to finance the cultural segment — *almost entirely* in the case of books; and
- a tiny segment of pieces not generated by market forces, as best illustrated by large paintings under public orders.

Contrary to what can be observed in audiovisual services today, there has not been an intermediate segment in the book and painting sectors — a sector with *de facto* commercial intent, but benefiting from public support in the name of culture.

Culture: ‘à la recherche d’une définition perdue’

In sharp contrast with the book and painting sectors, it has been assumed that audiovisual services in most countries over the last 50 years have been so intertwined with culture that the *entire* sector has needed protection from free market and trade disciplines in order to keep cultural audiovisuals alive. Compared with the assumptions operating in the book and painting sectors, such a change of perspective is so dramatic that it can only be explained by a converging set of mutually reinforcing reasons — cultural, economic and political. This then raises

⁴ It is beyond the scope of this paper to describe the efforts of the fifteenth century Flemish school to launch linen-based painting, 20 or 30 times less expensive than its wood-based equivalent, in order to enlarge the market of potential buyers of paintings (Harbison 1995).

⁵ For instance, art dealers stock paintings for which they expect (sooner or later) some demand from national museums in order to trade them with paintings demanded by foreign purchasers but subjected to export bans — the whole trade being of course the source of huge profits for the successful dealers.

the question whether these reasons are stable, or whether some components are undergoing profound change, opening the door to key reforms in commercial and domestic policies.

Tight links between audiovisual services and culture rely on a particular definition of culture. Canada best illustrates the tightness of this link, because of its special history, although similar arguments apply in almost all countries. On the one hand, culture becomes the messenger of domestic identity:

In Canadian books, magazines, songs, films, radio and television programs, we are able to see and understand ourselves. We develop a more cohesive society and a sense of pride in who we are as a people and a nation. (SAGIT 1999, p. 1)

Culture has lost its sense of ‘universalism’ (manifesting the reputation of a prince or an idea to the rest of the world). None of the five principles guiding Canada’s cultural policies evoke the role of Canadian culture in the rest of the world. The third principle only mentions a place for Canadian cultural products in the *Canadian* market (SAGIT 1999, p. 5). Such an inward-looking perspective certainly has some value — as the contingent valuation survey made in the case of Australia (Papandrea 1997) may suggest (though the survey also suggests that there is plenty of room for improvement). But such an inward-looking perspective also enables routinely uncreative sitcoms, made in a few hours with domestic directors, actors and studios, to be eligible for ‘culture’. On the other hand, because the Canadian focus on identity creates the risk of uniformity, cultural diversity constitutes one of the five principles guiding Canada’s cultural policies (SAGIT 1999, p. 5). Domestic identity and cultural diversity are inconsistent in many respects. But they work in the same direction in one case — both encourage protectionist policies that minimise competition from the outside and limit competition to ‘insiders’, such as the high barriers at the borders and massive domestic subsidies seen in the European Community.

Why the dramatic change in perspective? There is little doubt that the main driving force was the powerful European ‘nationalistic’ ideal of the nineteenth century, which progressively imposed an association between a ‘culture’ and a ‘country’ or its people — an association between groups, instead of interactions between individuals (artists and fiercely competing patrons or collectors, with successful artists playing sometimes the role of patrons or collectors).

Ongoing changes

This explanation raises two questions.

-
- How has this nineteenth century nationalistic view been able to survive for the last fifty years, despite the apparent relative antipathy to nationalistic notions over this period?
 - And why has this view so deeply permeated audiovisual services, and been so largely ignored in the book and painting sectors?

These two questions have a common answer. The nationalistic view would have had difficulties surviving the last fifty years if it had not been supported by certain economic and political features of the audiovisual industry. What makes the coming decade so exciting is that these economic features are vanishing — opening the door to fundamental changes.

The *economic* feature supporting the nationalistic approach in audiovisual services until the mid-1990s came from technical constraints and related market structures. Building a radio or television broadcasting network required massive initial investment that generated elements of ‘natural’ monopoly. The vertical integration of such networks and programming activities in often publicly-owned monopolies extended the natural monopoly elements to the whole audiovisual sector and, as television increasingly became the major outlet for movies and music, to the movie and music segments as well. Until the 1980s, movie studios and theatres also required large investments, amplifying the natural monopoly elements in the movie sector.

It is important to underline that these risks of natural monopoly and the possible existence of scale economies in audiovisuals do not occur before the distribution stage. Movies (bad or superb) are unique: shooting a higher number of movies does not decrease substantially their average cost of production. Scale economies occur at the distribution level when copies of a film are made in order to distribute the movie simultaneously in a large number of theatres. As a result, arguing the existence of scale economies for supporting public intervention in audiovisuals is incorrect.

- First, the argument should be limited to the distribution sector (more precisely, to movie copying), and not used for justifying subsidies at the production (shooting) stage. Moreover, the validity of the scale economies argument virtually disappears in the case of movies to be shown in TV networks, since scale economies are small in this case.
- Second, many modern industries show scale economies (and film distribution is definitely an industrial activity) without being subsidised. They cope with this feature by building appropriate industrial structures or mechanisms. Before asking for public support, one should thus look at whether existing regulations

do not inhibit the emergence of such structures (what follows suggests that it has been the case in Europe).

Similarly, scope economies are routinely invoked for public intervention in audiovisuals. As for scale economies, their mere existence is not enough to justify public intervention. It is a feature which can be handled by firms large enough to spread risks over a diversified portfolio of audiovisual works. Public intervention, by limiting the domestic market size, hence the emergence of large firms, reduces the ability to realise scope economies (ironically, public subsidies have increased the number of films without providing more scope economies).

The *political* feature underlying the inward-looking focus of culture and audiovisual services has been the ever-closer relationship between politicians and the audiovisual sector in increasingly media-intensive societies. Audiovisual directors and actors are increasingly seen as experts in ‘communicating’ with the wider society. Their influence becomes crucial at election times — then they are fixed assets in the political process and this position enables them to extract large rents from media-dependent political candidates. The fact that the audiovisual services industry is small (compared to the economy) makes it a perfect example of how vested interests work (Olson 1965). A small but powerful group rallies together to make a large financial gain by spreading the costs over the rest of society (implying small costs per person).

This argument may be seen as leaving aside other economic reasons for public intervention in audiovisual services. While it is beyond the scope of this paper to review all these reasons in depth, some observations are appropriate.

First, as noted, many of the arguments invoked as reasons for public intervention are actually consequences of, not cases for, public involvement. For instance, public intervention is said to be necessary because the audiovisual sector is prone to monopoly. This was correct in Europe, but it was because of public mismanagement that imposed artificial links between the physical audiovisual networks and content-related activities. Unbundling network access from programming activities would have allowed competition between content producers — indeed, this is progressively being done.

Second, some reasons for public involvement in audiovisual services are based on general arguments, but general arguments are not enough. For instance, it is argued that the audiovisual sector collectively produces social benefits (positive externalities) such as national identity, that it contributes to education, and that a free market would lead to under-production of audiovisual services. But why are audiovisual services more subsidised in the name of culture than are sports, presumably also a strong source of national identity? How can it be said that the

market will produce a sub-optimal number of films when many subsidised films are never shown, or are only shown for a few days? And even when general arguments seem appropriate (for instance, audiovisual services may cover a set of declining cost activities), there may be better policies for addressing the issue than public intervention (such as two-part tariffs). Once all these questions are raised and answered, the case for public intervention in the audiovisual sector as a whole does not appear very robust. Indeed, these arguments are also weak when applied to public support of culture (Heilbrun and Gray 1993), a point reinforced by the examples of books and paintings and re-examined the case of cultural audiovisual services (section 12.6).

The economic driving force behind the survival of the nationalistic model has been steadily eroding since the early 1980s. Technological change made radio broadcasting much cheaper in the late 1970s — widening the gap between program ‘organisers’ and program ‘producers’ by raising the issue of how to fill the growing radio broadcasting capacity with the more slowly increasing program supply. This erosion has just started, and will continue in the coming years. It is beyond the scope of this paper to describe in detail the extraordinary opportunities opened by new technologies and their capacities to expand massively consumers' choices (for a thorough analysis of the major opportunities, see PC 2000). One can simply mention that new digital television broadcasting technologies will allow a huge increase in the number of channels, multi-channelling by broadcasters, and a flourishing of ‘narrowcasting’ (cheaper and better technologies permitting a vast increase in ‘program differentiation’ capacities). In turn, this technological and economic evolution will profoundly shape the strategies of audiovisual firms, which will have increasingly strong incentives to become global in order to have more differentiated programs to consolidate. By the same token, large audiovisual firms may increasingly play the role of ‘patrons’ — not only of the culture of their country of origin, but also of the culture of their countries of destination — if this move enhances their reputation. A time could come where US firms do not hesitate to become patrons of young French moviemakers, and vice versa.

12.3 The protection of audiovisual services in the EC: ‘Mad tax’

In sharp contrast to books and paintings, audiovisuals in the European Community (EC) are subject to huge and arcane protection and to widely and tightly managed domestic markets. The EC Common Audiovisual Policies (CAPs) are the joint province of the Community (the Council of Ministers and the Commission) and the Member states. In other words, both levels of decision-making are required for defining regulatory reform and commercial policy in audiovisual services. As a

result, although the CAPs are based on similar instruments (hence justifying the ‘common’ qualification), they vary between Member states in their detail and style of enforcement (hence justifying the plural). The differences among Member states are not as significant as they first appear, and are subject to a slow ‘communitarisation’ process.

The CAPs instruments are close to those of the EC Common Agricultural Policy, suggesting that the consequences in audiovisuals will be similar to those in agriculture (as indeed is shown in section 12.4 below). However, there are two major differences between the Common Agricultural Policy and the CAPs. EC audiovisual services are dominated by a few large firms that are more aware than large and small farmers of the net costs of not becoming global. Second, these large EC audiovisual firms have markets of prime interest for them outside the European Community, namely in countries speaking the same language. These two differences are crucial enough to deserve a few remarks at the end of the section.

Tariffs and taxes

EC tariffs on films (for theatres) or cassettes (for individual consumers) range from exemptions to 6.5 per cent. There are no import quotas or bans, other than those for reasons of public order.

In sharp contrast, domestic taxes are substantial. In France, a tax of roughly 11 per cent is imposed on the price of theatre seats, independent of the nationality of the movie shown. As this tax finances subsidies granted to French films, it has a discriminatory impact, although to an extent that is relatively difficult to assess (see section 12.4 below). Special annual taxes specifically used to fund public television channels are imposed on television receivers in a handful of EC Member states. Lastly, there are taxes on blank cassette sales, on the grounds that such cassettes may be used to circumvent the many quantitative restrictions listed below, in particular the broadcast quotas. Such video taxes exist in various forms in almost all EC Member states.

Taxes imposed by sub-state entities play an important role in some Member states. In certain EC Member states (Germany, Belgium and Spain), the audiovisual policies are under the joint province of the Member state and its own sub-national entities. This generates three de facto levels of decision making in the European Community (the Community, the Member states, and the sub-national entities of certain EC Member states), requiring a complex legal framework, such as the 1991 German Broadcasting Treaty between the German Länder (States). The layer of sub-national entities involved can be very low. For instance, Belgian municipalities

have imposed taxes on television dishes, in order to limit the growth of satellite television channels competing with domestic channels.

Quantitative restrictions

Greece has quotas on first runs of movies in Athens and Thessaloniki. However, laws or decrees allowing such restrictions to be imposed also exist in France, Portugal and Spain ('protected' by GATT Article IV, see section 12.4).

The Television Without Frontiers (TWF) EC Directive (adopted in 1989 and revised in 1997) imposes 'broadcast quotas' based on film 'nationality'. A 'majority' of the total annual number of movies broadcast by every EC television channel should be of European origin. In France, this provision is interpreted as reserving a quota of 40 per cent to French movies, and a quota of 20 per cent to films originating from other EC Member states — leaving a maximum share of 40 per cent to non-European (de facto US) movies. In most of the other Member states, the proportion reserved to EC works is 51 per cent.

These quotas are rigidly enforced by certain Member states. In France, a lot of Conseil Supérieur de l'Audiovisuel (CSA) resources are devoted to counting every second of French, European and US works broadcast in France, to check whether the two reserved quotas (for French and EC audiovisual works) are being fulfilled. These provisions are implemented more subtly in other Member states (such as Britain and Germany), which use an incidental clause of the TWF Directive (quotas should be imposed 'where practicable') in order to introduce some degree of freedom to the constraint. However, this flexibility may be more apparent than real. The 'where practicable' clause has allowed the almost exclusive broadcasting of US films by the British channel BskyB. But relatively few US audiovisual works are broadcast by the British Broadcasting Corporation (BBC) and other British channels, so that in the end, the shares of US works broadcast in Britain and France are almost identical (and close to 40 per cent).

Broadcast quotas are combined with other quantitative restrictions in two Member states. France enforces the following four types of quotas, whereas Italy implements only type (iii) and (iv) quotas:

- (i) 'global' quotas limit the total number of films to be broadcast;
- (ii) time-specific quotas prohibit the broadcasting of films at certain hours and on certain days on 'air-television' channels [free-to-air television channels in Australia];

-
- (iii) arcane rules impose quota-equivalent constraints on the time ‘windows’ used for programming films in theatres, television channels and videos (these rules have been loosened recently for unsuccessful films); and
 - (iv) quotas on investments requiring every television channel to invest a share of its resources in film production by ‘pre-purchasing’ and/or ‘co-producing’ films: 3 per cent of the net turnover in the case of air-television channels, 20 per cent in the Canal Plus case (out of which 9 per cent is for French films).

The first three types of quotas systematically distort domestic competition — quotas (i) and (ii) protect movie theatres from television channels, whereas quota (iii) protects theatres and television channels from VCR and cassettes. Type (iv)-quotas are a consequence of restrictions on competition imposed on television channels (ie the pay-channel monopoly granted to Canal Plus). All these anticompetitive effects are amplified by the fact that the whole quota regime is defined on a firm-by-firm basis. Canal Plus can broadcast 364 movies, whereas the four major air-television channels can only broadcast from 170 to 192 movies per year, depending on the channel. Moreover, Canal Plus can broadcast as many films as it wants during ‘prime time’ hours, whereas other air-television channels can broadcast only 104 movies during these hours.

From an economic perspective, the key question is whether these quotas are binding. The answer is undoubtedly yes for the EC-wide broadcast quotas on US films, and the corollary is that such quotas are likely to be all the more harmful because they are introduced in a very imperfectly competitive environment. These quotas are strongly binding, since US films have a much higher than 50 per cent market share in all EC Member state cinemas (where EC viewers can make free choices) — from 63 per cent to more than 90 per cent (depends on the Member state). In fact, the binding effect of the quota may be perceived as more severe in Britain and Germany than in France, as the theatre-based market shares of US films are larger in these two Member states than in France.⁶

It is doubtful whether the other quotas are binding, with the exception of the investment quotas that have a powerful impact (see below). The French quota on the total number of films that can be broadcast annually is probably harmless, because the quota has been substantially increased over time. Now, the number of films broadcast in France is not so different from the number in EC Member states that do not enforce such a restriction. The same could be said for type (ii)-quotas. The type (iii) ‘window’ quotas are most costly for unsuccessful films (since they

⁶ However, the different market share in France may reflect the more extensive use of other protectionist tools (including subsidies).

prevent producers from reducing their losses by selling cassettes as quickly as possible) — although unsuccessful (but not successful) films are now exempted from such quotas.⁷

Subsidies

European subsidy regimes are very complicated in their funding sources and end uses. Table 12.1 focuses on the many sources of funding: grants from Member states and sub-state authorities, taxes on television advertising and on television receivers, and a wide range of other sources — from video taxes to National Lottery revenues (as in Britain).

Table 12.2 focuses on the different uses of these subsidies.⁸ In a nutshell, film subsidies can be divided into aids to production (column 3), which can be automatic (column 4) or selective (column 5), and other aids (column 2), a large part of which go to movie theatres. Aids to production can be given in the form of grants or ‘advances’ (hence raising the issue of the *effective* level of reimbursement, which is close to 100 per cent in Britain, but close to 15 per cent in France (CNC 1997)). It should be emphasised that aids to theatres are de facto highly discriminatory. They are officially justified on the basis that a large number of theatres are necessary for showing a wide range of new films, hence for generating the desired cultural diversity. However, this rationale does not fit reality — 20 per cent of French film production never reaches theatres, and most of the remaining new French movies are only shown for a few days. A much more plausible rationale for the survival of these large subsidies to movie theatres is that 50 per cent of the French theatre seats are owned by the French major studios (Gaumont, Pathé and UGC) which, being movie-makers, can shift these subsidies to production or to other uses.

⁷ In some Member states, such as France, quotas have been imposed on music since the mid-1990s, roughly on the same basis as broadcast quotas. The first years of implementation witnessed an increase of the share of ‘French’ music — a fact interpreted by quota supporters as evidence of quota efficiency. However, this success could be because the high level of competition prevailing before the quota implementation screened the best French musicians so that the quota just amplified the success of this existing stock of musicians. If so, the progressive erosion of competition under the quota regime will reduce the quality of French musicians, and progressively undermine the apparent quota efficiency. Indeed, this was the scenario observed in the film case. Last but not least, the so-called French music is simply the French translation of reggae, rap, etc, for the benefit of large music companies able to sell more expensive French and imported music.

⁸ A 318 page guide lists 17 broad types of subsidies applying only to feature films (excluding regional aids) (Fougea, Kalck and Rogard 1993).

Table 12.2 contains crude estimates of the subsidisation rates applying in 1995. Subsidies to production can be split between subsidies to film production (column 6) and subsidies to audiovisual works (the difference between columns 3 and 6, not shown). A first set of estimates of the subsidisation rate (column 7) is obtained by dividing the subsidies granted to film production by the revenues from theatre tickets to domestic films in each Member state. Subsidisation rates are large — the EC average is almost 40 per cent. Very high subsidisation rates for the small EC Member states flow from the fact that such countries do not grant large subsidies (in absolute terms), but they produce a few films screening to small audiences. However, the main estimates severely under-state the rates of subsidisation because they do not take into account subsidies not directly related to production that are, nevertheless, an integral part of the whole process. For example, as discussed above, France also subsidises theatres. In order to include this aspect, table 12.2 provides alternative estimates of the subsidisation rates based on the sum of the subsidies granted to film production and distribution. The average subsidisation rate for the EC film industry rises to more than 50 per cent, with huge increases in France and Germany, and little change in Britain.

Table 12.1 Funding sources of subsidies to EC audiovisual services
Million ECU, 1995

<i>EC member</i>	<i>Grants</i>		<i>Taxes and quasi-taxes</i>			<i>Other sources</i>	<i>Total subsidies</i>
	<i>State</i>	<i>Sub-state</i>	<i>non-TV</i>	<i>Public TV</i>	<i>Private TV</i>		
Austria	11.9	2.2		3.7		4.1	21.9
Belgium		23.8					23.8
Britain ^a	—	—	—	—	—	—	31.0
Denmark	26.1						26.1
Finland	8.7		2.7				11.5
France	56.3	5.9	88.1	77.4	140.0	3.9	371.6
Germany	8.2	61.3	31.0	39.4	7.4		147.2
Greece	3.8		0.5			0.8	5.2
Ireland	3.8						3.8
Italy	88.2		7.3				95.5
Luxembourg	1.5						1.5
Netherlands	9.9			23.7		1.0	34.6
Portugal	1.2			10.4			11.6
Spain	20.4	6.9					27.3
Sweden	13.9		9.6	1.6	0.6		25.7
EC-15	253.8	100.2	139.1	156.2	148.0	9.8	838.1
Per cent	30.3	12.0	16.6	18.6	17.7	1.2	100.0

— Breakdown not available. ^a Including funds from the National Lottery.

Source: Estimates based on European Audiovisual Observatory (1998, 1999).

Table 12.2 Use of subsidies and subsidisation rates in EC audiovisual services

1995

EC member	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	<i>Subsidies</i>		<i>Subsidies to production^b</i>			<i>Sub prod.</i>	<i>Subsidisation rates</i>	
	<i>Total</i>	<i>Non prod.</i>	<i>Total</i>	<i>Select.</i>	<i>Auto.</i>		<i>estimate 1</i>	<i>estimate 2</i>
	ECU m	% ^a	ECU m	%	%	ECU m	% ^c	% ^d
Austria	21.9	--	--	--	--	--	--	--
Belgium	23.8	44.0	13.3	71.7	28.4	11.1	1 078.3	1 078.3
Britain	31.0	9.3	28.1	100.0	0.0	9.2	8.8	8.8
Denmark	26.1	32.5	17.6	83.4	16.6	17.6	142.5	147.7
Finland	11.5	34.4	7.5	100.0	0.0	5.9	282.1	299.9
France	371.6	49.9	186.3	29.5	70.5	61.1	22.8	47.5
Germany	147.2	50.1	73.5	90.1	9.9	36.5	28.1	42.6
Greece	5.2	8.9	4.7	100.0	0.0	4.7	239.3	239.3
Ireland	3.8	0.0	3.8	100.0	0.0	3.6	390.0	390.0
Italy	95.5	4.3	91.3	92.0	8.0	91.3	70.6	71.3
Luxembourg	1.5	31.8	1.0	100.0	0.0	0.4	1 818.2	1 818.2
Netherlands	34.6	16.3	29.0	100.0	0.0	6.0	167.1	174.6
Portugal	11.6	50.5	5.7	68.4	31.6	5.5	652.7	695.3
Spain	27.3	21.2	21.5	51.8	48.2	17.7	38.2	42.3
Sweden	25.7	40.8	15.2	75.0	25.0	15.2	77.8	81.6
EC-15	838.1	37.9	498.6	66.2	33.8	285.8	39.7	52.2

ECU m: million European currency units. Non prod: non-production. Sub prod: subsidised production (films only). Select: selective. Auto: automatic. ^a Per cent of subsidies not directly allocated to production of films or television works. ^b Films and television works. ^c Subsidies granted to film production divided by the revenues from theatre tickets to domestic films in each Member state. ^d Sum of the subsidies granted to film production and distribution divided by the revenues from theatre tickets to domestic films in each Member state.

Source: Estimates based on European Audiovisual Observatory (1998, 1999).

Tables 12.1 and 12.2 do not give an idea of the evolution over time of the subsidisation process in EC audiovisual services. The following discussion focuses on France, because it is the Member state with the best data on subsidies and theatre ticket sales (these data are not strictly comparable with those of table 12.2, hence the discrepancies between table 12.2 results and the following results). In France, the estimated rate of subsidisation increased from 16 per cent in 1975 to 70 per cent in 1998 (71 per cent in 1995, instead of 47.5 per cent in table 12.2), if it is assumed that film revenues are almost exclusively drawn from theatre tickets. However, since the early 1990s, films have earned additional revenue from their release on television channels and from video revenues. This additional revenue was estimated at 25 per cent of the total revenues earned in the early 1990s (Farchy 1992). Taking into account these additional sources of revenue, the subsidisation rate has increased from 16 per cent in 1975 to 55 per cent in 1998 (52 per cent in 1995). But these

estimates are still based on a narrow definition of subsidies (as shown in table 12.2), because they ignore the indirect subsidies flowing from investment quotas. In particular, Canal Plus, with an average spending of ECU 130 million per year (FFR 800 million) in the late 1990s, contributes two-third of all funds flowing from investment quotas.⁹ Adding these indirect subsidies to direct state aid gives a subsidisation rate for French films close to 100 per cent in the late 1990s (as an explicit counterpart of its monopoly on pay television).¹⁰ Such a skyrocketing subsidisation rate raises the question of the economic impact of the CAPs.¹¹

12.4 Economic impact: ‘The castle of Count Dracula’¹²

This discussion of the EC CAPs raises two key questions.

- Have these huge transfers from taxpayers and viewers to audiovisual ‘producers’ been successful?
- Have they been particularly successful for the cultural segment of the audiovisual sector?

The answer is that not only have these efforts failed globally, but that their failure has been bigger for cultural audiovisual services. This has led an increasing number of culture-minded Europeans to look for regulatory reform.

⁹ The monopoly rent that Canal Plus can extract from its French television viewers can be estimated at ECU 300 million per year (FFR 2 billion) in the 1990s. This means that almost half of Canal Plus monopoly rents are a tax on French television viewers in order to subsidise French film producers. Since 1997, Canal Plus has had a competitor, TPS, which is a joint venture of all major French air-television channels (public channels and their archrival private channel, TF1), France Telecom and Lyonnaise des Eaux (the utility company, archrival of Vivendi). However, Canal Plus dominates pay television with 5.7 million subscribers compared to 0.6 million subscribers for TPS. Despite this, the rent erosion process has probably started.

¹⁰ This result may be perceived as an over-estimate because it ignores the successful movies. But this result also ignores a last source of subsidies — those paid to factors of production of films. Until two years ago, unemployment benefits for actors and other film workers were extraordinarily generous (even by French standards), with a budget of ECU 330 million per year, and were unfair (benefits were proportional to wages and fees). Subsidies for television satellites (TDF) have also been ignored.

¹¹ A similar escalation has almost certainly occurred in all the other EC Members.

¹² This section draws on Messerlin and Cocq (2000).

CAPs have been harmful to industrial audiovisuals

How to measure success? The number of films produced is not a good indicator, as it simply mirrors the fact that more public money is made available to filmmakers and spent by them. As, and when, EC Member states have progressively increased their subsidies (including under EC subsidy programs, such as 'Media'), their film industries have grown. Subsidies have encouraged the other large EC Member states to produce roughly the same number of films as France, the first to heavily subsidise its film industry.

A better indicator of success is the share of 'national' films in theatre ticket sales (the number of theatre tickets for 'national' films divided by the total number of theatre tickets) or, alternatively, the share of US films ('industrial' films receive a heavy weighting in both indicators). Until the mid-1990s, the domestic film share was much higher in France than in other EC Member states — leaving the impression that the French film policy was successful, compared to the less restrictive market policies of the other Member states. However, the late 1990s saw an unrelenting convergence of the domestic film share in France towards the shares in all other large EC Member states. This occurred despite arbitrary French decisions on rules of origin, such as the classification of the *Fifth Element*, a perfect Hollywood clone made by a French director, as a French film, which boosted the 1997 French film market share in France from 29 to 34 per cent.

Another interesting indicator is the share of non-domestic European films in each EC Member state (for instance, the share of non-French EC films viewed in France). This indicator gives a sense of the impact of the set of CAPs on EC audiovisuals as a whole. Not only is this share small, but it has generally been declining during the 1990s. This suggests that the CAPs may constitute a *barrier* to an integrated EC film market, contrary to their stated objective. This observation is not so surprising. National subsidies mechanically boost film production, and 'freeze' national investments in each Member state, while ignoring possible comparative advantages. But these subsidies are unable to create a demand for 'European' films. As a result, Member state subsidies have merely crowded out films from other EC Member states. Ironically, the French film industry has suffered the largest crowding out effect. And also ironically, the only EC film industry with growing EC market shares is the British film industry, which for years underwent severe cuts in subsidies (forcing it not to mimic US films, but to search for a British touch), and recently has seen noticeable investments from the US film industry.

CAPs have been even more harmful to cultural audiovisuals

There are two different — both admittedly crude — indicators that focus on cultural audiovisuals. Interestingly, each indicator corresponds to a different definition of culture.

The first indicator is based on television audience (hence, it seems closer to industrial than to cultural audiovisuals) (Cocq 2000). Despite heavy subsidisation, and hence a high level of production, there are not enough French films to feed the reserved quota for French films on television during prime time. Between 1994 and 1998, there were, on average, only 10 French films per year having more than 1 million viewers in theatres, compared to an average of 24 US films per year. Thus, few French films have been successful enough to be considered as candidates for prime time television. As a result, television channels have increasingly resorted to massive re-runs of old French films (the re-run rate has increased from less than 60 per cent in the 1980s to 70 per cent in the 1990s) or to made-for-television fiction works. This has relegated recent French films to feeding the non-peak hours of the day in order to meet the annual 40 per cent quota of French films. Ironically (but in many ways consistently), French films produced before the 1980s (hence with a substantially lower rate of subsidisation), show, even nowadays, a better capacity to attract viewers during prime time television, despite the handicap of many re-runs, than the heavily subsidised films produced during the 1980s and 1990s. Moreover, such a difference in the capacity to attract television viewers is not observed for recent US films.

The second indicator assumes that people who can recognise the talents of the French ‘culture’ — the ‘élite’ — roam the film Festivals. If this assumption is correct, the French ‘share of awards’ at the three major European film Festivals (Berlin, Cannes and Venice) seems a crude, but acceptable, indicator of the quality of French films. The French share of awards at these three festivals has decreased dramatically, from roughly 16 per cent (1981–1985) to 11 per cent (1986–1993) and 6 per cent (1994–1998). If Cannes is excluded, the corresponding shares are 21, 10 and 9 per cent, respectively.

All these data are consistent with basic lessons from the economics of trade policy. The huge increase in subsidies (in particular after the 1989 Plan) has fuelled production of high-cost French entertainment films ‘à la Hollywood’. Such French films provide large private profits if successful, but sizable public losses absorbed by the French Treasury if the films are a failure. More crucially from a ‘culture’ perspective, by specially protecting French ‘clone films’ from the comparative advantage of Hollywood, this policy has accelerated the ‘Americanisation’ of French industrial film production, by inducing French producers and directors to

turn away from cultural films and to focus on the secure and profitable niche of the market — films à la Hollywood. Moreover, the committees distributing the subsidies for cultural movies have also encouraged more homogeneous production instead of more diversification. The outcome seems the worst possible — an accelerated ‘Americanisation’ of French industrial film output and a sterilisation of French cultural film production.

Eroding support

The CAPs are increasingly showing symptoms of fatigue. Support for the CAPs has eroded significantly in recent years, from both the large audiovisual firms and the public. This discomfort was recently expressed in press reports on wasteful subsidies and interviews with television CEOs calling for reform (Messerlin and Cocq 2000).

However, reform may be difficult to initiate. The existing CAPs are so complex that many participants want the status quo, simply because they are unable to assess whether they will gain individually from reform, even though there may be a net gain overall. Table 12.3 illustrates these conflicting forces for the major participants: theatre-owners, film producers (US, EC, French successful and French unsuccessful), television channels and French consumers. Of course, an analysis in terms of effective rates of assistance (ERAs) would be of prime importance here (Plunkett, Wilson and Argy 1992).

12.5 The limitations of the GATS for audiovisual services

For audiovisual services, rules of origin, quantitative restrictions, subsidies and domestic regulations are the four major obstacles to trade relevant in a GATS context. This section focuses on two questions.

- To what extent could these trade barriers raise specific difficulties in handling the audiovisual sector in the coming WTO Round?
- To what extent do they hurt cultural audiovisuals substantially more than industrial audiovisuals?

The second question is particularly important because it indicates whether (and where) clarifying the definition of industrial and cultural audiovisual sectors will be beneficial to the point that it is likely to be supported by a coalition — in particular, by those concerned with culture.

Table 12.3 French audiovisual policy: 'Au nom de la culture'

	<i>Costs</i>	<i>Benefits</i>	<i>Rents from protection</i>
Theatres	seat tax	subsidies ^c broadcast quotas on TV	
Film-makers			
Hollywood	seat tax broadcast quotas on TV	seat tax ^d	yes, except the best
EC	seat tax broadcast quotas on TV	subsidies	no (unlikely to bind)
French I ^a	seat tax	subsidies	
French II ^b		subsidies	
Television channels			
Air channels	investment quotas, taxes		
Canal Plus	investment quotas, taxes		yes, from monopoly
Viewers	taxes, quotas, rents		

^a Successful French filmmakers ^b Unsuccessful French filmmakers ^c If not shifted to production. ^d Equipment and comfort.

Source: See text.

Before looking at these four major obstacles to trade, two preliminary remarks are necessary.

First, a vast majority of WTO signatories — among them, the EC — have not made any 'specific' commitment on audiovisual services within the GATS, which are therefore excluded from GATS disciplines of market access, national treatment, and additional commitments (Articles XVI to XVIII). This does not eliminate the fact that EC audiovisual services have to comply with the GATS 'general obligations and disciplines' (that is, GATT Articles II to XV). In particular, the Members have to comply with Article II (most favoured nation treatment, though the EC has made use of the MFN exemption provided for by Annex to GATS Article II in order to cover bilateral agreements on 'cultural' exchanges), Article V (economic integration — listing the conditions necessary to escape from most favoured nation (MFN) disciplines in the case of 'preferential' trade relations) and Article XV (on subsidies). WTO panels are possible on all the matters covered by these Articles.

Second, the history of the Uruguay Round negotiations provides a sense of the provisions considered crucial by the audiovisual vested interests in the mid-1990s. France (backed by Spain and Belgium) was in favour of a 'cultural exception'. Such an approach would have required adding an Article XIVter to GATS Article XIV (general exceptions) and Article XIVbis (security exceptions), excluding from the

scope of the GATS all categories of audiovisual services under all present and future forms of technologies. The European Commission (backed by Britain and Germany, among others) was in favour of a more limited ‘cultural specificity’ aimed at restricting the scope of three GATS provisions:

- the Annex to Article II on most favoured nation treatment (by adding a clause allowing discriminatory co-production agreements in the audiovisual sector to last more than ten years);
- Article XV on subsidies (by adding a ‘flexibility’ clause); and
- Article XIX (by excluding, in case of ‘cultural values’, the audiovisual sector from the principle of progressive liberalisation — a principle which, if it gives more time for certain WTO Members, keeps the final goal of always increasing their general level of commitments).

Both efforts failed. The key reason was, probably, the fear among European governments and businessmen that excluding audiovisuals would incur too high a price in terms of forever excluding other sectors in which the European Community had some comparative advantage.

These aborted attempts deserve two more remarks. First, they show how difficult it is to introduce an exemption of a ‘general’ nature (such as the French proposal) into the GATS framework. Despite its many weaknesses, the GATS has an intrinsic strength flowing from the global WTO framework. The weak point of the French proposal was that implementing a cultural exception would have risked WTO panels continually checking whether the exception really involved audiovisuals or not. That is, the French proposal risked leaving the definition of the scope of the exception to non-EC Member states. Second, the only instrument-specific exemption tabled by the European Commission was on subsidies. The fact that the EC proposals did not include Article XVI (quantitative restrictions) reveals the difficulties of building a coalition on this instrument, and sections 12.3 and 12.4 suggest that this conclusion will probably be increasingly robust in the future.

Rules of origin

There is no definition of, and no procedure on how to define rules of origin in the GATS text. While this problem applies to all services, it raises particular difficulties in the case of audiovisuals — often resulting in inconsistencies and farcical paradoxes.

Turning first to the inconsistencies, it is widely agreed that questioning the country of origin makes little sense when one talks about culture. Wondering whether Monteverdi is Cremonan, Mantouan or Venetian is clearly a silly question, and

stating that he is Italian adds anachronism to nonsense. If the history of world cultures sends a clear message, it is that artists and art works travel as much as possible in order to combine influences. Thus, it seems inconsistent to defend culture by imposing rules of origin so alien to it, but so beneficial to industrial vested interests (as trade economists know so well from the GATT and WTO history). Interestingly, there has been no attempt to apply rules of origin to ‘scientific culture’ — there is an instrument (patents) at the disposal of scientists in every country which makes the ‘origin’ issue irrelevant. Similarly, book authors, architects, etc are protected by copyrights rather than rules of origin.

Such inconsistencies are more than conceptual — they have the capacity to undermine the operation of the international trade regime. For instance, the United Nations Educational, Scientific and Cultural Organization (UNESCO) is busy expanding the concept of ‘world heritage’ in an effort that extracts culture from narrow political borders. Proposing to deal with audiovisual issues within UNESCO, without establishing beforehand a distinction between industrial and cultural audiovisuals, would shift the above inconsistencies from the international trade architecture to inside UNESCO.

Leaving aside this basic problem, rules of origin in audiovisuals are likely to be solved in the WTO forum by adopting rules about local content, following the trend in goods trade. Rules of origin have increasingly been used the last two decades because of the expanding number of regional trade agreements, but experience shows that their use is unsatisfactory. It further distorts production patterns between countries, makes trade facilitation procedures more costly and generates corruption in public administration.

These flaws of a rules of origin regime will be even bigger in the case of audiovisuals, and they will be particularly costly for the cultural segment of the audiovisual sector. This is because there is no way to define a culture by its inputs. For instance, stating that, say, a French director gives a French ‘passport’ to movies raises two problems. First, a film made by a French director can be a perfect clone of an ‘American’ film — the last 20 years of the French film policy has left plenty of examples, and anyone looking at French television is welcome to test this assessment. Second, a film made by a foreign director permeated with French culture would not be considered French, even though it may be the best trophy of French culture.

There are more technical problems. In sharp contrast with goods (for which local content can be captured in one global indicator, such as a change in tariff lines or a percentage of value added), there is no possible synthetic indicator in audiovisuals, as best illustrated by movies. Plenty of criteria could be used for defining the local content of a movie (the director, the actors, the initial story on which the movie is

based, the language used for the script or the location of shooting, etc), and, of course, there is no way of aggregating them in a meaningful way. The current French regime includes no less than 12 criteria, ranging from the language to the director, actors, scenery, equipment and other technical steps. This makes it a farcical illustration of the ‘inventaire à la Prévert’, if only because the same weight is granted to the actors as to the equipment. And yes, we are talking about culture!

The first way to solve the aggregation problem would be to require the fulfilment of *all* the listed criteria. This approach would be costly for industrial audiovisuals, but one could count on clever businessmen adjusting to this ‘one size for all’ policy. This is best illustrated by the movie *Le Affinità elettive*, based on Goethe’s work, directed by two Italians, shot in Tuscany and declared to be 100 per cent French on the grounds that 78 per cent of the dialogue was in French (three of the major actors were French).¹³ Culture-minded audiovisual providers are less likely to have such a flexibility — making the system more harmful for culture than for business. Of course, the more criteria on the list, the more likely it is to hurt the cultural segment of the audiovisual sector.

In order to overcome these difficulties, there is little doubt that (pushed by vested interests in audiovisuals) the WTO will go one step further on such a Kafkaesque path, and decide that different weights should be given to the various criteria listed, and that nationality should be granted on a threshold basis. For instance, language could be worth 20 points, domestic star actors 10 points, etc, and the movie would be ‘domestic’ if it got more than 80 points (out of a total of 100 points). Such a solution would create more problems than it solved. For instance, what about regional languages? (In France, regional languages are accepted, allowing German, Italian and Spanish on board.) Why should star actors receive only 10 points (out of 100) when the success of a movie relies largely on them? Or would this small share be convenient enough to allow ‘French’ movies to rely on the reputation of foreign actors? Of course, from a culture-minded perspective, all these problems — and their solutions — produce Frankensteins, and disadvantage culture-minded moviemakers compared to industrially-minded directors.¹⁴

¹³ Opportunities for cheating are as large as in the Common Agricultural Policy. For instance, it would be possible to shoot in English, and to post-synchronise in order to get the ‘preferred’ language. French authorities have threatened to test for this by showing films to deaf people (able to recognise the language by looking at the actors’ lips). You said ‘culture?’

¹⁴ And they increase the likelihood that a film can get several nationalities (independently from cases of joint production).

Quantitative restrictions

There are two different types of quantitative instruments relevant to audiovisual services:

- those imposed on services *per se*, either directly (the number of films) or indirectly (the time available for broadcasting foreign programs); and
- those imposed on factors of production, in particular limits on ‘foreign’ casting in ‘domestic’ films and the capital ownership of foreign investors.

This second source of quotas is particularly important in audiovisual services, which are characterised by extensive periods of working abroad. Relaxing or tightening constraints on such an input is likely to have a huge impact on the output, since a few actors are a recipe for success, or at least, for damage containment — a key advantage in the highly unpredictable world of audiovisual works.

Concerning quotas on services *per se*, audiovisual services are the only ones subjected to a GATT provision (Article IV) explicitly *authorising* quotas. Article IV allows WTO Members to impose screen quotas, defined as a percentage of the total screen time actually utilised. This provision echoes several agreements between the United States and European countries (Britain and France, among others).¹⁵ It is unlikely to be harmful in the near future, mostly because technical progress tends to shift audiences away from movie theatres, and because the US Majors have invested in movie theatres outside the United States (and many theatres would be unprofitable without Hollywood films). Moreover, the wording of GATT Article IV is precise enough to make it difficult to extend its use to other audiovisual works, although the GATT Working Party of the early 1960s unsuccessfully sought to extend Article IV to television programs.¹⁶

As a result, GATS Article XVI on market access is likely to play a much more important role. It prohibits quotas, be they based on the number of service suppliers,

¹⁵ The so-called *Blum-Byrnes Agreement*, signed in 1946 by the United States and France, was negotiated at the request of the United States Government. Before World War II, there was an annual quota on dubbed films in France (150 US films out of a total of 188 films). In late 1944, the post-Vichy French ‘*gouvernement provisoire*’ imposed a ban on all foreign films, arguing from the reduction of the number of movie theatres and from the elimination of the ‘double program’ regime (which allowed a spectator to see two movies with one ticket: its elimination was equivalent to a price increase).

¹⁶ Despite the absence of a formal agreement on extending GATT Article IV to television programs, the mere existence of Article IV represents a danger. For instance, there are regularly new pressures to re-impose screen quotas in France. However, as long as the US Majors retain their comparative advantage, the danger is small.

on transaction values or on asset values ‘unless otherwise specified in the country’s schedule’.

Lastly, concerning disciplines on factor-based quotas, the GATS has very little about the movement of labour and ownership, and it has nothing about investment, except about limitations on the participation of foreign capital (these are irrelevant for the EC, which has made no market access commitment). The lack of provisions on the movement of labour is likely to be more costly in the cultural segment of audiovisuals than in the industrial segment. The reverse applies for constraints on capital. As the movement of labour is more constrained than capital flows, the absence of disciplines in the GATS may be more harmful to cultural than to industrial audiovisuals.

In any case, the widespread ten-year exemption for audiovisuals makes it impossible to test the strength of Article XVI disciplines. As a result, the coming decade will witness crucial debates on the two following questions.

- In countries implementing protectionist audiovisual policies, to what extent will an emerging understanding that quotas have been inefficient (see below) induce these countries to take commitments under GATS Article XVI?
- Even more importantly, will there be a growing recognition that quotas have been a particularly harmful instrument for protecting cultural audiovisuals?

Subsidies and domestic regulations

These two issues should be examined together, as both deal with the same fundamental objective of designing and implementing the best possible domestic regulatory framework for audiovisuals in general, and for cultural audiovisuals in particular. It should be noted that regardless of international problems, subsidies raise domestic questions because they are transfers between nationals. As these transfers are costly to design and implement, they raise the question of whether lower subsidies would lead to a better cultural outcome through better design of domestic regulations. Because they are interconnected, subsidies and domestic regulations should constitute the core of a Reference Paper that is probably necessary for integrating audiovisual services into the WTO (see section 12.6).

GATS Article VI on domestic regulation is very general. Its limitations are not as severe as they may seem at first glance. GATS Article VI contains all the ‘seeds’ needed for drafting a Reference Paper because it goes further than the usual stance on transparency, which, although useful, is far from sufficient. In particular, GATS Article VI contains the two key elements needed for successful regulatory reform. It insists on the ‘lightness’ of the regulations (‘not more burdensome than necessary’),

which echoes cost minimisation for a given objective. And it ranks highly the two major qualities of suitable dispute settlement procedures (speed with ‘prompt review’, and action with ‘appropriate remedies’).

By contrast, GATS Article XV (subsidies) is an almost empty shell in terms of the disciplines it imposes on subsidies. The current draft is of little help in addressing the ease with which subsidies can be introduced on behalf of culture, and for industrial purposes. The only interesting aspect of GATS Article XV is that it does not consider countervailing procedures (which have been disastrous in the case of trade in goods) as necessarily ‘appropriate’. This possibly opens the door to better anti-subsidy disciplines, all the more essential in the case of audiovisuals. Because of the ambiguity of GATS Article XV, WTO rules on subsidies to be inserted in any Reference Paper will be difficult to draft. It will be even more difficult because subsidies are the instrument for which it is easiest to build a coalition between the industrial and cultural segments of audiovisual services.

12.6 Liberalisation and regulatory reform: ‘Independence day’?

The coming decade begins with better prospects for integrating audiovisuals into the WTO than existed under the Uruguay Round. Profound changes have occurred in the European Community and United States, generating a much less antagonistic world than six or seven years ago. However, the GATS framework could not become fully operational without the adoption of a Reference Paper laying down certain obligations specific to the audiovisual sector. In particular, the paper should make clear the distinction between industrial and cultural audiovisuals, in order to exempt cultural audiovisuals from certain WTO disciplines. This section describes key changes that would be propitious for the coming negotiations, then raises a few essential issues that the Reference Paper should address and concludes by insisting that liberalisation would only be fully beneficial if domestic regulatory reforms were also adopted.

A rapidly changing world

In contrast with the general acceptance of CAPs in the European Community until the mid-1990s (Messerlin 1997), an increasing number of people are realising that they are inappropriate for ‘promoting culture’, instead often harming culture (see section 12.4). Moreover, most large EC audiovisual firms are changing their mind. The technical evolution in audiovisuals (cable, satellites, and telecommunications) has seen the integration of some audiovisual companies into parent companies with

many other business lines. As a result, reciprocity is now a concept working *inside* EC firms. If Vivendi (the parent company of Canal Plus) or Suez Lyonnaise des Eaux (the parent company of M6, the only successful French air-television channel) want open access to US utility markets or to US radio waves, they have to be ready to accept liberalisation of EC audiovisual markets. This evolution has been accelerated by the realisation that ‘large’ EC firms are small relative to their American (North and South) and Asian counterparts. This ‘smallness’ is largely a consequence of the market segmentation generated by the CAPs and it constitutes a heavy handicap — less because it limits scale economies than because it limits risk spreading, making the long-term survival of these firms doubtful in an audiovisual industry characterised by huge and unpredictable risks.

In the United States, Hollywood still enjoys a unique capacity to produce movies that can be seen all over the world with equal success. It is a major European misconception that Hollywood produces ‘American’ movies — instead, it produces ‘worldwide’ movies. But US Majors are facing rising costs at home, forcing them to invest outside the United States and, hence, to become closer partners of audiovisual firms in the host countries. At the other end of the audiovisual spectrum, ‘narrowcasting’ has opened niches in which non-US firms have been quite successful (often by learning from and cooperating with US firms). Entry into US markets becomes a realistic target for these firms — a modest start to two-way trade in audiovisual content. Lastly, internet-based technologies reduce audiovisual entry costs, an evolution that will favour non-US firms (which tend to have a comparative advantage in ‘non-mass’ markets) more than US firms (which tend to have a comparative advantage in ‘mass’ films and have invested in theatres accordingly).

A last important change is the erosion of the EC coalition supporting highly protectionist CAPs (all other things being equal), largely because EC Member states face very different situations in terms of the prevalence of their language, hence market size and structure. Three Member states (Britain, Spain and Portugal) have large markets with the same language outside the European Community. This makes it easier for viewers from these Member states to perceive benefits from foreign competition, especially since powerful non-EC audiovisual firms operate in their non-EC markets. Three EC Member states (France, Germany and Italy) have relatively large markets with the same language in Europe, but limited markets outside the European Community. While Germany and Italy are accustomed to this situation (they have always had limited territories outside Europe), France is not. As a result, a strong coalition of these three countries is unlikely (as shown during the Uruguay negotiations). A strong coalition is even less likely given that the key French audiovisual firms are fast trying to internationalise their activities, and have been included in parent companies with worldwide perspectives (best illustrated by

the position of Canal Plus after the proposed merger of Vivendi and Seagram). Lastly, the other Member states are accustomed to two or more languages, making them relatively open (maybe with the exception of Belgium).

The need for a Reference Paper

The GATS limitations (section 12.5) and the resulting weakness of the existing legal framework make complete liberalisation difficult without the support of a Reference Paper, especially if the problem of culture is to be addressed (as it should). The following discussion presents a few key issues to be addressed from a European Community perspective.

GATS Article XVI (market access) offers enough generality and precision to cope with all the problems raised by a progressive relaxation of typical quotas. In particular, GATS Article XVI offers a robust basis for negotiating the elimination of broadcast quotas (additional type (i) to (iii)-quotas do not really matter, see section 12.3). If EC governments are strong enough (and confident enough in European creativity), quotas reserved for domestic audiovisual works could be eliminated in a few years. If EC governments are faint-hearted and invoke the GATS concept of ‘progressive’ liberalisation, the minimum achievement in the coming Round would be a bound reduction in quotas to the level of the share of domestic films in theatres. This would reduce the reserved quota in France from 60 per cent to roughly 30 per cent, and from 51 per cent to 20 per cent or less in other EC Member states.

However, GATS Article XVI has some difficulties in handling the reduction and elimination of investment quotas. GATS Article XVI is written in general enough terms to allow such a reduction or elimination. It could even allow a reduction on a firm-by-firm basis. For instance, constraints on Canal Plus could be reduced more rapidly than the investment obligations imposed on the other firms. But this evolution raises a question that GATS Article XVI cannot address. Canal Plus’s specific investment obligations flow from its monopoly power, and that power should be eroded at the same pace as its obligations. In other words, the full use of GATS Article XV on subsidies may require the use of Article VI on domestic regulations (in addition to market access). This is where a Reference Paper becomes necessary.

In particular, a Reference Paper should offer specific ways for handling progressive liberalisation of the audiovisual subsidy schemes, under the constraint that cultural audiovisuals — and only these — should not be limited by certain WTO disciplines (eg subsidies), though they should be subject to other WTO disciplines (eg the no quota commitment). This goal requires that the Reference Paper provide, not a

definition of culture (an impossible task), but proxies for culture that would be accepted worldwide and could be used for bargaining. The following discussion provides some tentative illustrations.

- A first proxy could be the number of films made by a director. The first film of a director, say, could be largely subsidised, with the subsidies being reduced for the second film, and eliminated for the third and subsequent films. Such a proxy aims at addressing several goals:
 - it fits the frequent request for cultural diversity;
 - it seems consistent with the fact that a director able to create two films successfully (with declining subsidies) is likely to find the funds necessary for additional films;
 - it is unlikely to involve massive subsidies (for instance, the subsidies granted to ‘first’ films in France represent only 5 per cent of total annual French subsidies); and
 - it allows some progressivity to initiate the process.

Of course, such a proxy can be circumvented (if only by the use of figureheads). As a result, some additional constraints could be imposed, such as a cap on the subsidy. By definition, a subsidy on a first film should be small since the budget of such films is small. For instance, the average subsidy for first films in France is roughly US\$ 300 000.

- Such a proxy could seem inadequate to certain WTO Members, because it relies on the assumption that ‘cultural’ movies by confirmed directors could not be costly (indeed, film history suggests that it is the case). An additional approach would be to negotiate an ‘amber’ box — a maximum amount of annual subsidies, or a maximum number of films to be subsidised — in addition to the first movie(s) exemption. However, it is clear that such an amber box could easily generate the same protectionist drifts as those shown by the blue box in agriculture.

All these disciplines aim to eliminate the amount of subsidies granted to industrial audiovisuals and to make sure that subsidies granted to cultural audiovisuals will not be diverted to industrial audiovisuals. However, such liberalisation will only be fully beneficial to culture if accompanied by appropriate domestic regulatory reform.

Domestic regulatory reforms

Eliminating industrial subsidies and allowing cultural subsidies is possible because these two segments of the audiovisual sector are quite distinct (Cocq 2000). This

point is well illustrated by a recent and bitter dispute between French film directors and critics that revealed three categories of directors: those making successful industrial films; those making unsuccessful industrial films; and those making cultural films (Cocq 2000). Pressure for additional protection came from the second group, but it was striking that they got no support from the first and third groups of directors.

However, if a Reference Paper is a necessary condition for improving support for culture, it is certainly not a sufficient condition. Domestic regulatory reforms are also required.

Democratic regimes can hardly nurture culture if they do not recognise and correctly handle the key aspect of culture — that it has to be a ‘bet’. There is almost no way of knowing in advance which bets will go into the treasuries of human kind, which ones will be important preparatory works, and which ones will be deadends or wastes. Risk is thus the key factor. Hence, the absolute necessity of ‘real’ patrons, that is, individuals or small groups of people who risk their money and/or their reputation. Venetian doges, French kings, Chinese emperors and American billionaires have all taken *individual* risks by asking certain architects, painters and writers to work for them. They have spent their money on, and linked their fame to, a painting, a musical composition or a literary piece. In fact, there were complex and competing pyramids of risks, with aristocrats and commoners ‘testing’ new styles under fierce competition. This is, perhaps, best illustrated by Louis XIV using the same team to build Versailles as his Finance Minister (who ended his life in jail) did to conceive and build his castle of Vaux-le-Vicomte (often called the ‘blueprint’ of Versailles).

Democratic governments cannot exercise this ‘patron’ function in a direct way. When they try to do so, they create bureaucratic committees that tend to overspend, preferably on friends and in a reciprocal way, and/or tend to spread the money in a politically correct way (leaving France today more ‘bourgeoisie’ and ‘conventionnelle’ in many artistic respects than it was under Louis-Philippe in the 1840s). This inability of democratic governments is a well-known problem (indeed, hotly debated during the French Revolution (Monnier 1995)). Thus, there is a need to look at the best institutions that could, in a democratic regime, nurture culture. The key challenge is how to design such ‘patron-institutions’ in order to make them as interested in cultural creativity as possible (and as different as possible from the existing bureaucratic committees). Without meeting such a challenge (which goes beyond this paper), liberalisation would simply be neutral for culture. In fact, liberalisation may be beneficial for culture only to the extent that it improves the way artists are funded.

There are many candidate institutions that could improve funding for artists. The following discussion briefly mentions two aspects relevant to audiovisuals in an international context. First, culture sheds some unusual light on intellectual property rights. If it is important that such rights exist and are enforced, the current tendency to expand these rights is worrisome. Longer copyrights (from 50 years to 70 years in the European Community) or stronger enforcement (as with the recent WTO panel on music rights, or the payment of royalties to the architect of the French Très Grande Bibliothèque in return for taking its photograph) could generate ‘excessive’ costs that could endanger the emergence of future culture. This brief observation may suggest that any Reference Paper could also look at the Uruguay Agreement on Trade-related Aspects of Intellectual Property (TRIPs) in the context of culture, possibly by integrating work done in the context of the Treaties regulating art works, and should be careful to balance the TRIPS agreement with appropriate competition policies.

Second, subsidies allowed by any Reference Paper should be better used. Some useful conclusions might be drawn from comparisons between the various EC schemes for granting subsidies — from the completely centralised and Byzantine French system to the more decentralised, but no more successful, German regime, to the relatively more successful (so far) British ‘franchise’ framework. The British approach is the only one that tries to maintain competition in ‘patrons’ production’ — capturing the core aspect of what is a patron. Improving this aspect may be as important for culture as focusing on reforming funding sources (shifting from a public to private funding¹⁷) or reforming distribution structures (public television channels or specific narrowcasting channels could show ‘cultural’ audiovisual works to a wider audience than theatres, without the risk of being crowded out by ‘industrial’ movies). The fact that ‘cultural’ audiovisual works would benefit much more from e-commerce than ‘industrial’ works in terms of audience and reputation should be taken into account by the European Community when looking at the e-commerce issues. It should induce the European Community to be more open on e-commerce issues than the United States.

References

CNC (Centre National de la Cinématographie) 1997, *La Rentabilité des Films*, Paris, October.

¹⁷ Although, it could be argued that setting up competition among patrons to spend public money may lose an important motive to do well, namely to lose money.

-
- Cocq, E. 2000, *Analyse économique de la politique française du cinéma*, PhD thesis, Institut d'Etudes Politiques de Paris, Paris, unpublished.
- European Audiovisual Observatory 1998, *Comparative Analysis of National Aid Mechanisms*, vol. 1, Council of Europe, Strasbourg.
- 1999, *Statistical Yearbook*, Council of Europe, Strasbourg.
- Farchy, J. 1992, *Le Cinéma Déchaîné*, Presses du CNRS, Paris.
- Fougea, J.P., Kalck, A. and Rogard, P. 1993, *Les Aides au Financement*, Editions Dixit, Paris.
- Harbison, C. 1995, *The Art of Northern Renaissance*, Calmann and King, London.
- Heilbrun, J. and Gray, C.M. 1993, *The Economics of Art and Culture: An American Perspective*, Cambridge University Press, Cambridge.
- Messerlin, P.A. 1997, *European Film Policy: 'La Grande Illusion'*, Centre for Policy Studies, London.
- and Cocq, E. 2000, *Preparing Negotiations in Services: EC Audiovisuals in the Millennium Round*, American Enterprise Institute and Groupe d'Economie Mondiale, Washington and Paris (forthcoming).
- Monnier, G. 1995, *L'art et ses Institutions en France de la Révolution à Nos Jours*, Collection Folio-Histoire, Gallimard, Paris.
- Olson, M. 1965, 'The Logic of Collective Action: Public Goods and the Theory of Groups', *Harvard Economic Studies*, vol. CXXIV, Harvard University Press, Cambridge.
- Papandrea, F. 1997, *Cultural Regulation of Australian Television Programs*, Bureau of Transport and Communications Economics Occasional Paper no. 114, AGPS, Canberra.
- Plunkett H., Wilson R. and Argy, S. 1992, 'The measurement of effective rates of assistance in Australia', OECD Working Paper ESD/DIV(92)1, OECD, Paris.
- PC (Productivity Commission) 2000, *Broadcasting*, Report no. 11, AusInfo, Canberra.
- SAGIT (Canadian Cultural Industries Sectoral Advisory Group on International Trade) 1999, *Canadian Culture in a Global World: New Strategies for Culture and Trade*, SAGIT, Ottawa.
- WTO (World Trade Organization) 1998, 'Audiovisual Services, Background Note by the Secretariat', Sectoral Paper S/C/W/40, Council for Trade in Services, WTO, Geneva.

Discussant — *Pierre Sauvé*

Harvard University

This paper provides an insight into the structure of the EC audiovisual regulatory regime. As a Canadian, it would be foolish to argue that this sector is immune from wasteful rent-seeking and the kinds of excesses that Patrick Messerlin's paper usefully deplores.

Having said that, the paper is profoundly French. The incestuous relationship between a country's cultural elites and the political system is not as prevalent elsewhere. In Canada, there is only one officially registered intellectual, John Ralston Saul, and few people make a living from being social critics or essayists. Yet in France they abound, and are courted by the political system. This has a profound bearing on how French policy is shaped.

There is much to agree with in Patrick Messerlin's paper. I come at the topic from a Canadian perspective, which Patrick would bemoan as being parochial and inward-looking. But a look at the data on Canada's openness to foreign cultural goods and services, particularly American goods and services, shows that there is more than mere inward-looking parochialism in Canada's quest for policies to provide some measure of cultural diversity in markets and space on shelves for domestic content.

It is easy for trade economists to dismiss these arguments as subtle forms of protectionism disguised under elegant cloaks.

I would very much disagree with this characterisation. Patrick Messerlin states that he values the universality of culture, but this is a profoundly French viewpoint. The analogies drawn with the book and painting industries in 15th century Europe are not altogether convincing or relevant to this debate. There may have been some segmentation between industrial books and cultural books. But the size of this trade, the kind of people it reached and the impact that it had on collective identity was of an entirely different nature to that of media and audiovisual products today. The analogy is an interesting one, but I would not deduce the same kinds of policy conclusion that he does.

The discussion of quotas in the paper is interesting and highlights their design flaws and limitations. Most of these support measures are technologically challenged. Perhaps wasting negotiating effort on policies that will be undermined by technology is not a good use of time. From a perspective of efficient regulation, these measures impose compliance costs that probably exceed the potential efficiency gains. So these measures are soon to be dinosaurs.

That said, quantitative restrictions and quotas do not have the same effects in all market segments. They have tended to work better, for some reason, in the music industry than in audiovisuals. Certainly, sales of French and Canadian artists have picked up significantly since Canada and France enacted quantitative restrictions on music broadcasting. Perhaps there is a need to compare the dynamics of these different market segments.

From a Canadian perspective, what I find interesting is that recent cultural disputes between Canada and the United States have involved a bargain. This has allowed greater foreign ownership by American producers of media content in Canada, in return for an acceptance by those foreign investors of obligations to provide local content, to abide by performance requirements, to make space for Canadian western music in broadcasting (in a celebrated case), or to include more domestic content in US-based split-run magazines.

So one possible approach to multilateral bargaining would be to open up the ownership restrictions that abound in these sectors, while accompanying this investment liberalisation with modest local content performance requirements. Clearly, this is not a free trade ideal, but is likely to be a bargain worth pursuing.

My final point is on subsidies. Patrick Messerlin makes a strong plea to progressively eliminate all industrial subsidies, while keeping subsidies for cultural products. I am a bit sceptical about this recommendation. In the paper, the distinction between the industrial and cultural component of the industry is not made clear. What criteria would distinguish industrial from cultural? The assumption seems to be that cultural products are those that fail, while industrial products are those that either succeed or fail, but the distinction is not clear.

From a multilateral perspective, why are subsidies a concern? Do they have demonstrably injurious effects on foreign providers in the marketplace? If so, perhaps there is a case for developing disciplines, but I do doubt if this is the case. European or Canadian subsidy programs, which in a way are the most efficient forms of protection imaginable, are unlikely to have such distorting effects as to injure third country producers.

So my recommendation is that subsidies should be a matter for domestic political debate, rather than for international rule making. Individual countries will determine through the political system whether or not they wish to subsidise cultural industries.

General discussion

This is reported at the end of the following chapter.

13 Towards multilateral rules on trade and culture: protective regulation or efficient protection?

Pierre Sauvé and Karsten Steinfatt[†]

Harvard University and Organization of American States

13.1 Introduction

While many ‘trade and ...’ problems have only recently achieved notoriety in policy circles (Trachtman 1998), conflicts between trade and cultural values go back to the 1920s, when a number of European countries introduced quotas in order to protect their fledgling film industry from a sudden influx of films from the United States. Despite its long-standing nature, the ‘trade and culture’ quandary is by no means resolved. On the contrary, against a global backdrop characterised by the increased commoditisation and homogenisation of mass-consumption cultural products, countries around the world have expressed an increasing desire to protect their national identity, values and beliefs through a range of policies on culture. Not surprisingly, given the high tradability of many cultural products (goods and services), the last few years have witnessed escalating friction between trading nations over the place of such products in international law. Such frictions have brought to the fore the shortcomings of the existing institutional framework in properly balancing trade and cultural objectives.

Unless something is done in the coming years to more resolutely address such shortcomings, chances are, particularly in the context of the World Trade

[†] The authors would like to thank Professor Joel Trachtman for his comments and suggestions. Parts of this paper were first presented by Karsten Steinfatt at Professor Trachtman’s Seminar on the Legal Aspects of International Economic Integration at the Fletcher School of Law and Diplomacy at Tufts University. Seminar participants, in particular Gastón Bilder, Elizabeth Cadena and Shigehiko Nishizawa, also provided helpful comments. Special thanks are also due to Greg McGuire and participants at this conference for their helpful reactions to the paper. The views expressed in this paper should not be ascribed to the Organization for Economic Co-operation and Development, the Organization of American States and their respective member countries.

Organization's (WTO) environment of significantly heightened judicial activism, that the existing tension between trade and culture will increase rather than decrease (Bernier 2000 and Carmody 1999). To date, attempts to resolve that tension have centred on proposals designed chiefly to shield cultural goods and services from the application of international trade disciplines, whether through cultural exceptions, exemptions, reservations or other similar legal devices (Bernier 1998). Such an approach is unlikely to prove sustainable as a long-term policy option. For one, as the rising number of cases recently submitted to WTO dispute resolution attest, cultural products are already largely subject to WTO disciplines governing trade and investment in goods, services and trade-related intellectual property matters. From a juridical standpoint, any purported cultural 'exception' or blanket immunity from international trade law, a notion still much in vogue in large segments of some countries' cultural industries, is mere chimera.¹ Second, ongoing technological developments and the proliferation and convergence in media platforms that they allow are blurring already murky boundaries between goods, services and IPR-related issues in the cultural field. Recent WTO panel and, especially, Appellate Body decisions, most notably those relating to periodicals² and bananas,³ reinforce the above point.

The question of whether WTO disciplines should apply to cultural products is a moot point. However, a far more relevant question is to determine the kinds of disciplines and the delicate blend of adjudication and legislation that is most likely to strike an acceptable balance between the competing objectives of cultural diversity and market openness.

This paper focuses on two dimensions of the trade and culture debate.

¹ However, it bears noting that except for Articles IV and XX(f) of GATT, there is no direct invocation of domestic cultural policy issues anywhere in WTO law. Article IV (Special Provisions Relating to Cinematograph Films) allows member countries to maintain 'screen quotas' and specifies a number of requirements to which quotas must conform. In essence, Article IV seeks to limit the application of screen quotas to 'films of national origin' in order to ensure the consistency of these measures with the most favoured nation principle established in Article I. Article XX(f), on the other hand, allows member countries to adopt or enforce measures 'imposed for the protection of national treasures of artistic, historic or archaeological value,' subject to the requirements set forth in the article's *chapeau* [preface].

² See Appellate Body Report: Canada — Certain Measures Concerning Periodicals, WT/DS31/AB/R (30 June 1997); and Panel Report: Canada — Certain Measures Concerning Periodicals, WT/DS31/R (14 March 1997).

³ See Appellate Body Report: European Communities — Regime for the Importation, Sale and Distribution of Bananas WT/DS27/AB/R (25 September 1997); and Panel Reports: European Communities — Regime for the Importation, Sale and Distribution of Bananas, complaints by Ecuador (WT/DS27/R/ECU), Guatemala (WT/DS27/R/GTM), Honduras (WT/DS27/R/HND), Mexico (WT/DS27/R/MEX) and the United States (WT/DS27/R/USA).

First, it asks whether the insights of economic theory can be deployed in such a manner as to challenge the normative assumption that the benefits of freeing up trade in cultural goods and services are just as compelling as for any other category of products. Simply put, can economics help advocates of cultural diversity reclaim the dominant public policy discourse, and restore some degree of equivalence between this discourse and the political (and negotiating) reality of the trade-culture interface? Arguing that, over and above their commercial, commoditised features, cultural products play a central role in shaping domestic democratic discourse, the analysis put forward in the paper lends support to the claim that the special features of markets in cultural goods and services can cause markets to fail to provide consumers with appropriate production and distribution of cultural products. Such features include:

- the nature of competition in products with substantial public goods aspects;
- economies of scale in the production and distribution of cultural goods and services;
- the impact of externalities on the pricing of cultural products; and
- collective action problems.

The broad tendency for markets in cultural goods and services to fail to yield welfare-maximising outcomes justifies recourse to corrective governmental interventions that may, among other things, take the form of subsidies and local content rules. Analysis of the economic problems free trade creates for potentially highly valued local cultural products, combined with the scope for implementing policies of ‘efficient protection’ (Sykes 2000) that minimise adverse effects on trade and investment, suggest that cultural trade policies existing in much of the world’s democracies, and the commensurate regulatory precaution shown to date by a great majority of WTO Members, are broadly sensible from both a political and economic point of view. That said, the paper argues that care should be taken in designing trade- and investment-restricting policies with a view to ensuring that foreign cultural content is allowed to play a useful contributive role in shaping domestic cultural dialogue.

Second, drawing on the insights offered by the law and economics literature on the issue of rules versus standards, the paper assesses whether the multilateral trading system’s current predominant reliance on adjudication in matters of trade and culture is a sustainable proposition in light of the growing incidence of commercial conflict arising from this policy interface. The analysis suggests that WTO Members may need to contemplate a shift from standards to rules, and to lay the foundations for a comprehensive body of multilateral disciplines governing trade, investment and IPR-matters in the area of cultural industries, aimed at reconciling the twin pursuit of market openness and cultural diversity.

13.2 Trade and culture: a closer look at the underlying economics

The United States has long sought to see the cross-border exchange of cultural goods and services governed by free trade principles. Its determination to codify this position in international trade law almost derailed the concluding phase of the Uruguay Round. Despite intense diplomatic and commercial efforts, the US view ultimately failed to carry the day. For example, to date, only a distinct minority (18) of the WTO's 136-strong membership has agreed to schedule bound liberalisation commitments in audiovisual services under the General Agreement on Trade in Services (GATS). This suggests that most countries deem it preferable, as a matter of trade policy, to refrain from taking on legally binding GATS obligations in order to pursue their cultural policies.⁴

Although the United States stands virtually alone among major trading nations in its negotiating stance on trade in cultural products, it is remarkable how the American position is widely seen, particularly among large segments of the country's economics and legal professions concerned with matters of trade, as winning all the theoretical arguments, especially rational economic arguments. Cloaked in freedom of expression arguments that command powerful intuitive and symbolic appeal, the mainstream view regards free trade in cultural products as justified as a matter of principle, and not simply as a manifestation of mercantile self-interest. The dichotomy between the political (and negotiating) reality of trade and culture, which reveals a clear majority preference for regulatory autonomy on the part of WTO Members, and the (largely unchallenged) dominant, pro-free trade, economic discourse directed to the interface, is striking.

The broad outlines of the competing paradigms in the trade and culture debate can be described succinctly. Widely accepted premises (and two centuries of practice) justify the general claim that free trade benefits people of all countries. Despite its complex distributional consequences, both over time and between and within countries and categories of workers, free trade is, by and large, widely understood as being good for consumers, as well as for a society's more productive workers and its most efficient firms. This premise underlies the WTO system and most of the world's attempts to progressively liberalise trade and investment regimes in recent decades.

⁴ These countries are: Central African Republic, Dominican Republic, El Salvador, Gambia, Hong Kong, India, Israel, Kenya, Korea, Lesotho, Malaysia, Mexico, New Zealand, Nicaragua, Panama, Singapore, Thailand, and the United States. Of the group, only two WTO Members, the Central African Republic and the United States, undertook market access and national treatment commitments in all six sub-categories of audiovisual services covered by the GATS (WTO 1998, p. 7).

The United States contends that such a premise applies straightforwardly to cultural products, especially films and television programs, but also to magazines, books, musical recordings and advertising, which collectively rank atop the country's sources of export earnings. Under this view, nothing economically special about these products is seen as justifying a repudiation of free trade premises. Deviations from free trade are seen as particularly harmful for consumers in countries imposing import restrictions, as they deprive them of the cultural products that they want as sovereign individuals.

An alternative view affirms the 'specialness' of culture as a federating force in pluralistic democracies (Baker 2000). Such a view draws attention to the fact that, over and above their commercial dimension, cultural products are importantly vehicles of social communication that promote a wider sense of belonging and cultural identity. Cultural patrimony can thus be seen as a crucial component of the identity and self-understanding of a nation. Accordingly, countries should have the right to protect and promote their own culture internally (box 13.1).

The dominant view, particularly in American trade policy circles, is to label this perspective of culture as 'paternalistic' and 'protectionist'. Even when accepting that cultural trade policy involves difficult choices between the free trade premise of letting people have products they want and the value of preserving local culture, commentators typically disparage the latter claim. They are prone indeed to argue that protectionists invoke culture as a smoke screen for purely self-interested policies designed primarily to benefit politically powerful domestic commercial interests.

There are admittedly few reasons to believe that public officials should be any more immune from the politics of capture in matters of cultural trade policy than with respect to any other sensitive policy interface. And there is assuredly some truth to the assertion that paternalism and elitism, together with industrial policy activism and socially wasteful rent-seeking, typically lurk in the second-best world in which cultural trade policy is formulated. That said, proponents of cultural diversity and of rules aimed at providing some immunity to governmental measures taken in support of domestic cultural content (most of whom have not to date emanated from trade policy circles) have generally done a poor job of questioning the soundness of the economic arguments aligned in support of the free trade premise on trade and culture.

Box 13.1 Clashing over the meaning of culture

The trade and culture debate tends to oppose two starkly differing views of culture. Free trade proponents often take what Baker (2000) has described as an ‘artefact’ or ‘commodity’ conception of culture, viewing it as content-oriented and backward-looking. Cultural integrity is seen as involving an essentially static preservation of this historical content from the perturbing effects of new (and especially) foreign ideas and influences. Free trade obviously threatens this preservation function. Its advocates focus on the political economy of who benefits from, and who pays for, preservation, often with the implicit suggestion that certain elites benefit while the broader public, who are denied the imported cultural products they desire, pay. Proponents of free trade in cultural products bemoan the parochialism in content deriving from protectionist policies and posit that free trade would allow domestic cultural industries to reveal their true value to the world and induce them to supply more popular, and hence more desirable, products. The latter suggestion assumes that the true value of culture lies in the universal appeal of its content rather than in its appeal to local audiences.

This conception of culture can be contrasted to one where participants, rather than content, occupy centre-stage — the primary audience being other members of the community. Rather than preserving backward-looking patrimony, the relevant goal of such a ‘culture as dialogue’ approach is essentially to ensure ‘shelf space’ and provide resources to members of a cultural community so as to foster dialogue on domestically salient issues (Schwanen 1998, Bernier 2000). Members’ speech to each other will often be contextually specific, even if universal meaning can sometimes be found in particular cultural output. So that even if free trade enabled a nation to export more cultural material, if it also caused the nation’s cultural industry to change its content to compete better in international markets but to speak less directly to domestic audiences, free trade could impoverish the domestic cultural dialogue.

The policy implications of this latter view of culture differ markedly from those suggested by the ‘commoditised’ view depicted above. If the overriding policy goal is to maintain or create a dynamic local cultural discourse, this cannot be done without preserving (or indeed creating) local cultural industries. Such a goal could justify direct subsidies as well as indirect means of support — for example, screen or broadcast quotas or content rules aimed at guaranteeing space for domestic producers and products, as well as some discriminatory burdens on imports. However, since the value of culture lies to an important degree in its ability to inform democratic choices, like censorship, any categorical exclusion of imports will tend to stunt discourse and is thus unacceptable. Outside cultural content and influences cannot be excluded since they usefully contribute to domestic dialogues on national identity and help shape a country’s social and political aspirations. Rising calls to embrace the broader notion of cultural diversity, and growing disaffection with the inward-looking and exclusionist logic of cultural exceptions, can be more readily understood in this context, as the recent shift in the international policy debate involves more than semantic subtleties (Bernier and Sauvé 1999).

As it happens, a credible case can be made to support the claim that the special features of markets in cultural goods and services can cause markets to fail to provide consumers with appropriate production and distribution of cultural products. Such features include the nature of competition in products with substantial public goods aspects; economies of scale in the production and distribution of cultural goods and services; the impact of externalities on the pricing of cultural products; as well as collective action problems. Because each of these problems arises in the context of international trade, there are strong grounds to believe that free trade in cultural goods and services is unlikely to yield efficient or welfare-maximising outcomes.

Thus, the challenge for policymakers becomes one of devising a hierarchy of domestic support measures able to remedy market failures at source. In doing so, the remedies should, wherever possible, minimise any adverse effects on trade and investment. The aim is to preserve the considerable benefits — the most important of which is the promotion of a better understanding among peoples — that accrue from the cross-border exchange of cultural goods and services. Stated simply, the challenge is to ensure that legitimate regulatory interventions represent, in the light of their likely high incidence on foreign products, ‘efficient’ forms of trade and investment protection.

Cultural products as public goods

Cultural products, and intellectual property more generally, share an important public good characteristic — their use is substantially non-rivalrous.⁵ This arises from the low cost (which is close to zero in the case of broadcast media) of providing them to additional users. Non-rivalrous use has both explanatory and policy relevance for international trade. For starters, media firms sell their products at markedly different prices from country to country.⁶ For most private goods sold in competitive markets, a substantially lower price for foreign sales strongly suggests ‘dumping’, that is, selling at below cost. Non-rivalrous use, however, means that the exported product is not sold at below cost because there are no significant additional costs beyond those already incurred by producing it for the domestic market.

⁵ Non-rivalry means that the consumption of a good or service by one person does not reduce the amount available for consumption by others (eg defence).

⁶ Shao (1995, p. 121), for example, quotes a mid-1980s survey conducted by the magazine *Variety* which found that prices for feature-length US films for television ranged from US\$90–US\$150 in Bermuda to US\$30 000–US\$40 000 in France. Similarly, the average cinema ticket price in the United States in 1993 was around US\$7, while it was US\$3.50 in Mexico.

This public good aspect of media products explains the readily observed dominance of the United States in international trade in films and television programming. Indeed, if the price to a consumer is the same, the customer will typically prefer the product that costs more to produce, that is, uses better inputs. Consumers seldom get this choice, for it is rare that goods that cost a lot to produce — the average cost of a Hollywood film was US\$52.7 million in 1998, as opposed to an average of less than US\$2 million in most countries with small film industries (Footer and Graber 2000) — sell for the same price as cheaply produced goods. The tradeoff for most private goods is typically between higher quality and lower price. Not so for media products, for as long as copy costs are minimal, the production budget for a single item will depend not on the amount one individual will pay, but on the number of paying customers to whom the firm can sell the single product (or its cheap copies). Putting aside international trade, a larger production budget should be available for media products primarily targeted at bigger and wealthier markets. Such a situation has long favoured large production budgets for American producers.

Enter international trade. If distributors sell domestic and imported media products at the same price, consumers in all countries are likely to prefer the products with the higher budget. More often than not, that will be an American product. Admittedly, most audiences will have a preference for domestic content — that is, content spoken in their own language and which addresses their own cultural and social predicaments. Predictably (if much to the ire of film distributors in Continental Europe who bemoan the closed nature of the US market),⁷ Americans will logically consume primarily home-made media products. For countries with smaller audiences over which to spread costs, the US advantage of higher budgets and the local advantage of domestic content should push in opposite directions, leading to some balance between imports and domestic products. Generally, the proportion of domestic products should increase, the larger and richer the domestic audience is and the more distinct its domestic culture and language are compared to those of exporting countries. Still, as the large bilateral imbalances reported for audiovisual trade between the United States and most OECD countries attest, the larger production budgets and the big and wealthy domestic audiences give American media products a significant edge in export markets.

Note how the greater ability of American media producers to tap export markets, based once more on the natural advantages flowing from the larger budgets afforded to them by the superior size and wealth of the domestic market, affects matters of

⁷ European films accounted for 3 per cent of box-office receipts in the United States in 1998, a year during which US films accounted for 63.5 per cent of box office receipts in Europe (Courrier de la Planète 2000, pp. 52–3).

content. It does so by tilting production away from products with predominantly domestic cultural content towards products with greater universal appeal. Although art critics typically praise the former types of cultural products, free trade may actually discourage their production. Seen this way, the oft-criticised ‘shallowness’ of American cinema or television production, with its predominant reliance on action, violence, sex and special effects, may owe less to domestic tastes or the lesser innate talent or imagination of American producers than to the industry’s significantly greater dependence on export markets. Hence, the predictive assertion that American media products will feature greater doses of products with ‘universal’ appeal than will cultural products emanating from countries relying less on export revenues. In the process, American exports will usually take market share from and reduce demand for an importing country’s domestic cultural products, while also crowding out American products with lesser export appeal.

Because producers of cultural goods and services are unable to price discriminate adequately, many cultural products will not be commercially valuable even though their creation and distribution would be a valuable use of social resources.⁸ This market failure is generally more common for cultural products with comparatively smaller audiences. In both domestic and foreign markets, media products with culturally-specific content fail even when strongly valued by local audiences and even though they would exist in greater quantities if free trade were subject to appropriate restrictions. In such circumstances, free trade may well hurt consumers.⁹

Economies of scale

The concept of ‘economies of scale’ refers to the tendency for unit costs to be lower with larger output. Economic theory holds that scale economies provide countries with a twofold incentive: first, to shift resources towards the production of a restricted number of goods (with a view to increasing output and reducing unit costs); and second, to engage in international trade, which creates an integrated market larger than any one country’s market (Krugman and Obstfeld 1997). When scale economies are present, the incentive for countries to specialise and to engage in international trade translates not only into a greater variety of products available for consumers but also lower prices.

⁸ Selling at high prices would cause the loss of too many customers to be profitable, while selling at a low price would not produce enough revenue per customer.

⁹ Free trade policies afford large budget productions greater capacity to price discriminate domestic productions with a smaller-audience. This inability of valued, yet small scale, domestic production to price discrimination and recoup production costs over various market segments (movie theatres, video-cassettes, television, cable, book spin-offs) may cause them to fail.

The fact that so many cultural industries, including film, television, music, and publishing, are characterised by economies of scale presents policymakers with the difficult challenge of balancing cultural and trade values. While free trade resulting from economies of scale will usually lead to efficiency gains in the form of greater variety and lower prices, in the case of cultural products characterised by scale economies policymakers cannot simply formulate and implement rules that promote free trade. As discussed above, economies of scale provide countries with the twofold incentive to trade and to specialise in the production of a limited range of goods. For some countries, specialisation might entail the loss of a cultural industry that is strongly valued by society. Policymakers might therefore wish to formulate ‘efficient’ forms of protection that limit the degree to which specialisation, in the context of cultural goods characterised by scale economies, occurs.

So far, our discussion has focused on internal scale economies, that is, those occurring at the level of the individual firm. However, not all scale economies apply at this level. Often, the concentration of production in one or a few locations may reduce an entire industry’s costs, making a cluster of firms more efficient than any single firm in isolation. In this case, we speak of external economies of scale.

Economic models of external scale economies and international trade assign an important role to history and accident in determining the pattern of international trade. Specifically, three sources of external economies are usually cited in the literature (Krugman and Obstfeld 1997):

- the ability of a cluster of small firms to support specialised suppliers;
- the creation of a pooled market for skilled labour around the cluster of firms; and
- knowledge spillovers.

Due to the pivotal role played by history and accident in generating and shaping each of the sources underlying external scale economies, it is not uncommon to find situations in which a country starting with a large industry retains its advantage in the production of a particular good even if another country could potentially produce the same good more efficiently.¹⁰ Again, this possibility might warrant the implementation of ‘efficient’ forms of protection aimed at containing or reversing certain patterns of trade that pose a threat to domestic cultural industries.

¹⁰ See Donaldson (1996) for a useful overview of the events that resulted in today’s trade pattern in audiovisuals.

Externalities

Markets achieve efficient outcomes when products are sold at their real cost. Improper pricing can occur for various reasons, including dumping — selling a product in the importing country below its cost of production in the exporting country, usually as a result of subsidy or subsidy-like practices directed at the exporting firm. Improper pricing can also occur when products create large negative or positive externalities — that is, costs or benefits to people other than their owner or purchaser. A product's real cost is the sum of the cost to producers as well as the cost imposed on third parties. Because negative externalities are costs not borne by either producer or purchaser, producers of goods with negative externalities can sell them at below their real cost.

Much as they are difficult to identify and measure, cultural products generate significant externalities, both positive and negative. By affecting readers' or viewers' behaviour, behaviour that in turn affects other people (why else would the advertising industry exist?), the cultural product has positive or negative value for people other than its buyer or seller. The existence of such externalities, however, does not *per se* justify measures that discriminate against foreign cultural products. For this to be the case, the externalities of imports would need to differ in a predictable (and negative) enough manner from those of domestic products that the imports would compete with or replace. Whether such differences exist will largely be context- or country-specific, and attempts at evaluating them will inevitably prove contentious.

The comparative success of imports that results from their larger production expenditures has in itself no *prima facie* relation to positive or negative externalities. Rather, it is the predictable differences in content flowing from the public good and economies of scale characteristics noted earlier that are likely to be most relevant for an externality analysis. Foreign cultural goods and services characteristically supply new and different perspectives. This can support greater cross-cultural understanding and bring with it notions of equality, liberty, democracy, the politics of race or gender, and other potentially 'foreign' ideas.¹¹ On the whole, therefore, measures aimed at fostering cultural diversity are likely to generate potentially significant positive externalities, to which public policy (including trade and investment policy) needs to be sensitive.

¹¹ These transformative properties can be especially important for individuals or groups oppressed or marginalised by a dominant local culture. It is hardly surprising that authoritarian regimes are deeply suspicious of free trade in cultural products and prone to varied forms of cultural censorship. The couching (and strong political appeal) of the trade-culture debate in freedom of speech terms is more readily understood in this light.

The introduction of transformative content through cross-border trade and investment may not always involve positive externalities. Imports may also have more negative consequences (for instance in the case of violent media) and indeed displace and crowd out local cultural products that have more positive aspects. For most importing countries, the dominant position of the US audiovisual industry means that foreign cultural products — and the greater degree of ‘universal’ content they typically feature — will often displace (rather than supplement) local products with high levels of domestic-specific content. The loss of domestic content because of displacement or marginalisation of local cultural products can generate potentially significant negative externalities, particularly from a ‘culture-as-domestic-dialogue’ perspective.

The quality and availability of news media, for instance, whether in print or broadcast form, performs vital democratic functions in any society, helping shape the development, awareness and richness of national identities. All these functions supply important positive externalities and can significantly improve the well-being of a community beyond the immediate consumers of such services. For such positive effects to be felt, domestic content and cultural specificity is typically crucial. In addition to using the local language, domestic-oriented cultural products typically reflect, interpret or render critical judgement on the predominant views found within the country from which they emanate. As Baker (2000) usefully observes:

To the extent that a country’s own media products better provide (or provide in ways not duplicated by imports) domestic content that people need directly for their political process to function or, more generally, for their cultural discourses of identity, meaning and value, domestic media can have tremendous positive externalities not supplied by the imports that threaten to replace the local content.

What policy implications can one adduce from the preceding discussion? For one, as with the case of market failures stemming from the public good dimension of cultural products and the presence of significant economies of scale in production and distribution, an externality analysis suggests that unrestrained trade in cultural goods and services may well reduce consumer welfare in the importing country. This result may obtain to the extent that the improper pricing of cultural products prevents audiences from getting what they most likely want — domestic content.

This market failure once again points to the desirability of corrective governmental intervention. The more imports of cultural products are seen to threaten or marginalise a country’s capacity to sustain a healthy dialogue on salient domestic issues, the greater the justification for enacting measures that impose ‘efficient’ burdens on imports or foreign investments. Accordingly, while shying away from any categorical exclusion of foreign cultural content that, in virtually all circumstances (including those giving rise to negative externalities) nourishes

democratic debates, countries should think twice (and typically do) before negotiating away their prerogative to impose some restrictions on trade and investment in cultural products. Rather, they should seek to codify international rules governing possible conflicts over trade and culture and agree (subject to dispute settlement disciplines) on an explicit hierarchy of admissible and least burdensome means of supporting domestic cultural industries under WTO law. Given the political sensitivities and the might of the industrial interests involved, doing so would assuredly represent a delicate balancing act. Indeed, because of its fluid nature, a concept like cultural diversity could very easily become the favourite veil used by certain groups to disguise protective measures. In turn, requiring the international trading system to indulge clearly protectionist policies on the grounds that they aim to promote domestic culture would dilute the considerable economic advantages derived from the WTO and the pursuit of free trade ideals.¹²

Collective action problems

In addition to the three sources of market failure depicted above, collective action problems might also warrant the formulation and implementation of policies aimed at protecting or promoting domestic culture. In his classic study on the interaction between groups and public goods, Olson (1965) showed that rational, self-interested individuals will not act to achieve their common or group interests, unless there is coercion or some other special device to make individuals act in their common interest. The reason for the inadequate level of collective activity can be traced back to what Olson refers to as ‘the essential’ function of a group or organisation, namely to provide ‘an inseparable, generalised benefit,’(p. 15) or public good. In most cases, there is a gap between the interest of a group to provide a public good and the individual interests of the members of the group.¹³

To illustrate how the inadequate level of collective activity operates in the specific context of cultural goods and services, suppose individuals in a community view

¹² This latter concern arguably informed much of the cautious reasoning displayed by first instance panellists and by Appellate Body Members in the recent dispute over magazines.

¹³ Olson draws a useful analogy between this gap and the gap between firms and industry in a competitive market: ‘[t]he firms in a perfectly competitive industry, for example, have a common interest in a higher price for the industry’s product. Since a uniform price must prevail in such a market, a firm cannot expect a higher price for itself unless all of the other firms in the industry also have this higher price. But a firm in a competitive market also has an interest in selling as much as it can, until the cost of producing another unit exceeds the price of that unit. In this there is no common interest; each firm’s interest is directly opposed to that of every other firm, for the more other firms sell, the lower the price and income for any given firm. In short, while all firms have a common interest in a higher price, they have antagonistic interests where output is concerned.’ (Olson 1965, p. 9).

domestic films as an essential manifestation of their culture. Accordingly, all individuals in this community would benefit significantly from an increase in the production of domestic films. One possible way to achieve such an increase would be if all individuals chose to view domestic films rather than films produced in other countries. From the perspective of the single individual, however, seeing only domestic films is unlikely to influence the capacity of the domestic film industry to increase its output. In fact, unless all individuals in a community decide not to see foreign films, a single individual's decision in this regard will have a negligible effect on the local film industry. In addition, for a single individual, the decision to limit his or her choice to domestic films might entail high costs, especially if he or she derives benefit from seeing foreign films as well. Meanwhile, any potential cultural benefits derived from the individual's 'sacrifice' would be available to all members of the community. In light of this collective action problem, government intervention, for example in the form of screen quotas for domestically produced films, might once again provide an avenue for a society to satisfy its preference for cultural diversity.

Preliminary conclusions

The preceding discussion suggests that economic theory cannot predict that free trade in cultural products will be better at giving audiences what they want than would a set of policies imposing some restrictive effect on imports. Indeed, when viewed from a 'culture as dialogue' perspective, economic theory suggests compelling reasons to expect free trade to be inefficient compared to an optimal system of trade restraints from the perspective of *both* American and importing country audiences. The most obvious policy response would be to subsidise those categories of cultural products most likely to be economically justified, but which will not survive in a free market environment. The US response to such policies has been to portray them as violations of free trade principles, even as it has failed to make concrete proposals on whether or how to curtail the use of trade- and investment-distorting subsidies in services trade. It has also tended to call for any such subsidies to be applied on a national treatment basis, even as the economic justification for subsidies (as is evidenced by the dominant market share of US media products in the vast majority of foreign markets) applies uniquely to local media products disadvantaged by the free trade regime.

In addressing the market failures identified above, a large number of countries have also enacted content quotas, typically requiring theatres, television networks, radio stations and publications to show or broadcast a certain percentage of domestically produced films, television programs, or music recordings or to feature local 'stories' or news events. Much disparaged by trade economists (and indeed condemned by WTO law in goods trade), such rules have nonetheless proven beneficial in some

instances as they marginally reduce the reach of imports without barring any particular import,¹⁴ even while giving rise to wasteful rent-seeking activity that may, in some instances, raise questions of distributional fairness, produce content of questionable quality and cultural specificity, and encourage anticompetitive conduct on the part of protected ‘infants’.¹⁵

By allowing imports to help make a theatre or radio station profitable, such rules can be regarded as a form of indirect subsidy to local content. While directly aiding local media products, properly designed quotas still allow importing firms to determine what combination of universal appeal and more culturally-specific imports will capture the largest remaining surplus. US dominance in large segments of cultural trade, particularly audiovisual trade, is such that a ‘remaining’ surplus will be quite large in most countries.¹⁶

¹⁴ Broadcast quotas for French language music in France, for instance, have had a clearly beneficial effect on sales of domestically produced music, which increased from 42 per cent of total sales in 1994, the year the quota policy was implemented, to 55 per cent in 1999, a rise of some 31 per cent. During the same period, investment in France by the largest recording companies expanded five-fold, from FFr 100 million in 1994 to FFr 500 million last year. What is more, by boosting the album sales of established artists, the policy allowed production companies to invest more in younger artists, to which content quotas also applied (Rony 2000, pp. 49–50).

¹⁵ Such instances call for greater competition policy activism, and a reconsideration of the antitrust immunity afforded in some countries to domestic cultural industries. A progressive relaxation of investment restrictions, even while maintaining — or indeed perhaps enacting — various local content requirements could also help foster greater competition in domestic cultural industries. The resolution of a number of commercial disputes between Canada and the United States relating to cultural matters (eg over broadcast rights for country music and, most recently, over magazines), typically featured such a combination of investment regime liberalisation and local content rules.

¹⁶ Canada, whose ‘protectionist’ cultural policies are constantly decried by US industry and documented in the Office of the US Trade Representative’s annual inventory of foreign trade barriers, vividly illustrates this point. In 1999, foreign (overwhelmingly American) content accounted for 45 per cent of book sales; 83 per cent of English-language consumer magazines on newsstands and 65 per cent of magazine circulation revenue; 79 per cent of the retail sales of tapes, CDs, concerts, merchandise and sheet music; 60 per cent of English-language television programming; and 97 per cent of screen time in Canadian cinemas. Similar, if slightly less spectacular, results obtain elsewhere, as the share of film receipts garnered by US industry stood in 1996 at 70 per cent in Europe, 83 per cent in Latin America and 50 per cent in Japan (Wernick 2000, pp. 52–3).

Box 13.2 Domestic support measures used in cultural industries

- *Subsidies*, involving the provision of grants and loans for the production of cultural works, most notably audiovisual products. For example, Eurimages, an initiative by the Council of Europe, provides subsidies for the co-production of European audiovisual works. The Media II program of the European Communities, while excluding the support of production, focuses on training for professionals, the development of attractive projects and the transnational distribution of audiovisual programs and films. National programs providing subsidies to the domestic film industry exist in France, Germany, the United Kingdom, Canada, the United States and Switzerland. Canada also subsidises its publishing industry through grants, marketing assistance, interest-free loans and postal subsidies.
- *Domestic content rules*, especially measures regulating radio and television broadcasting content. For example, the European Communities, the Council of Europe, Australia, Canada, and France use domestic broadcast content to control access to their television broadcast and film markets.
- *Market access restrictions*, in particular measures that control access to film markets, including screen quotas for cinemas (as in France, Mexico, South Korea and Spain); rebates on box office taxes for cinemas that show national films (Italy); prohibitions of dubbing of foreign films (Mexico); dubbing licences (eg in Spain, film distributors can only receive a dubbing licence for foreign films when they contract to distribute a certain number of national films).
- *Regulatory/licensing restrictions*, especially measures that control access to radio or television broadcasting through regulatory or licensing restrictions. Canada relies on this type of measure.
- *Tax measures*. France, for example, uses taxes on box office revenues, on receipts of broadcasters and on videocassettes to support local film production. Switzerland has been debating plans to introduce a tax on films that simultaneously are exhibited on more than 50 screens within the country.
- *Foreign investment and ownership* covers measures restricting foreign investment and ownership, including divestiture policies, for example, in the broadcasting industry and news media in Australia, Canada, the United Kingdom and the United States.
- *Border measures*, which may include tariffs or quantitative restrictions, as was the case in India, which used to restrict the import of film titles to 100 per year.
- *Film co-production agreements*. An example is the Council of Europe that has established a convention on plurilateral film co-production.

Source: Adapted from Footer and Graber (2000).

As a policy matter, screen and time quotas are preferable (and can indeed be viewed as a more ‘efficient form’ of protection) to quotas governing the *number* of imported cultural products. The latter will tend to favour imports of universal appeal ‘blockbuster’ products that are likely to prove more damaging to local industry (and domestic cultural dialogue) than would imports of smaller-audience products.

The foregoing analysis of the economic problems free trade creates for potentially more valued local cultural products, combined with the scope for implementing policies of ‘efficient protection’ that minimise adverse effects on trade and investment, suggests that the cultural trade policies that exist in much of the world’s democracies, and the commensurate regulatory precaution shown to date by a great majority of WTO Members, are broadly sensible from both a political and economic point of view.

The very real possibility that markets for cultural goods and services may fail to maximise consumer welfare under conditions of free trade suggests that policymakers should not set jurisdictional rules only with a view to enhancing market access conditions. They should also consider other factors related to the overlap between trade and cultural values.

A quick survey of the main types of cultural policies and instruments currently in place in different jurisdictions is indicative of the central role played by the various sources of market failures identified in this paper. Box 13.2 provides an illustrative list of the most commonly found support measures directed at domestic cultural industries. Debates over the WTO-compatibility of such measures and the possible need to codify an explicit hierarchy of permissible (ie non-actionable) and prohibited practices are likely to loom large in discussions aimed at developing international disciplines governing the trade and culture interface.

13.3 Legislation or adjudication? A law and economics perspective on how to manage a fractious policy interface

If, as is reasonably certain, the trade and culture interface will by its very nature generate commercial tension, an important question is how best to manage the policy interface. In considering these matters of governance and institutional choice, the law and economics literature provides useful tools of analysis.

In this literature, one essential distinction between rules and standards lies in the extent to which efforts to give content to the law are undertaken *before* or *after* the conduct regulated by the law takes place (Kaplow 1992, p. 560). A law is more

rule-like ‘the greater the extent to which it provides advance guidance’ (Kaplow 2000, p. 509). In contrast, a law is more standard-like to the extent that it establishes less specific guidance in advance of the conduct regulated by the law. From a rules and standards perspective, more specific law in the form of a rule is not always better law. Under certain circumstances, it might be more efficient to have only general guidance in the form of a standard prior to the occurrence of the regulated conduct, and delegate the task of giving content to the law *ex post* to a panel or tribunal. To identify the particular circumstances under which it might be more appropriate to have less specific law (ie standards) as opposed to more specific law (ie rules) in advance, the literature on rules and standards identifies several cost-benefit parameters.

- Costs of formulation: given that rules embody a higher degree of legal precision *ex ante*, they also entail higher costs in terms of formulation than standards. Such costs result from drafting, negotiation, and use of general administrative resources and will be directly related to the variability of the conduct that the rule or standard seeks to regulate. If the conduct in question has a large number of variants, it will be more costly to formulate a rule than a standard, for rules in this case will require those responsible for formulation to identify and define all possible variants of the relevant conduct and then decide on the scope and coverage of the law.
- Public choice costs: some commentators have observed that rule-making, particularly in the international trade context, is likely to be subjected to intense domestic scrutiny, while the application of a standard by a dispute resolution process is likely to be subjected to less scrutiny. According to this perspective, then, rules would entail higher costs than standards in public choice terms. It remains to be seen how the coming of age of internet-based policy advocacy by non-governmental organisations will affect the public choice costs and benefits of rules and standards.
- Primary and secondary predicability: rules are generally associated with a higher degree of primary predicability. For States as well as private parties, the interpretation of rules will usually entail lower costs than the interpretation of standards. Precisely because rules embody more information *ex ante*, individuals or States will not need to invest resources in assessing whether their behaviour conforms to the requirements of the law. Rules will be less expensive than standards in terms of secondary predicability as well, because rules usually allow parties, after the relevant conduct has taken place, to predict the outcome of dispute resolution and by implication, to resolve a dispute in a less costly manner.
- Legitimacy: since rules often involve *ex ante* specification through a treaty negotiated by elected government officials, rules are likely to reflect the

preferences of citizens to a greater extent than standards, which often involve *ex post* specification by panels or tribunals. Especially at the international level, the process whereby a panel or tribunal gives content to the law *ex post* is likely to be perceived as removed from citizens' preferences, particularly due to the lack of representativeness embodied in an international panel's composition.

Before using the insights provided by the rules and standards literature in addressing the question of which institution should manage the interface between trade and cultural values, we need to make explicit the institutional dimension of rules and standards. Rules often involve *ex ante* specification through treaties or political agreements, while standards often involve *ex post* specification by panels or tribunals.¹⁷ While this distinction might not always hold in practice (courts may be able to formulate rules pursuant to statutory or constitutional authority), the international trading system provides an example of a sphere where, generally speaking, the source of rules is likely to be a treaty or political agreement, whereas standards will be applied by a panel or tribunal. Viewed from this perspective, and disregarding the relative nature of the institutional dimension of rules and standards, the choice between rules and standards implies a decision about allocation of jurisdiction between dispute resolution bodies on the one hand, and legislatures on the other.

The law and economics literature on rules and standards assigns a central role to the 'frequency' of the conduct regulated by the law: all other things being equal, rules (or legislation) tend to be preferable when the frequency of the conduct to be regulated is high, while standards (or adjudication) are more efficient when the relevant conduct is relatively rare (Kaplow 2000, p. 510). The reasons can be traced back to the cost and benefit parameters outlined above. Rules have higher formulation costs (including public choice formulation costs) that are only borne once. On the other hand, the benefits of rules in terms of primary and secondary predictability and perceived legitimacy will arise often if the conduct required or proscribed by the law is frequent and rarely if it is not.¹⁸

Applying this insight to the domain of trade and culture suggests that as the frequency of 'trade and culture' conflicts increases, the net benefits of giving

¹⁷ This distinction is drawn from Trachtman (1999, p. 353).

¹⁸ To illustrate this point, consider the case of a law regulating the release into the water supply of hazardous substances. For dry cleaning and automotive fluids, it would be best to specify permitted disposal techniques in advance, as many small establishments would be unlikely to be able to determine the effect on their own and, even if they could, the costs of each obtaining information on the matter would be great. Redundant costs would also be incurred in inevitable enforcement proceedings. But for rare substances, it may not be worthwhile to specify the range of permitted and prohibited disposal techniques in advance, because most scenarios involving potentially hazardous action may never arise. The example is drawn from Kaplow (2000, p. 511).

content to the law *ex ante* through legislation will exceed the net benefits of giving content to the law *ex post* through adjudication. While this result is relatively simple to understand from a conceptual perspective, it might prove difficult to apply it to particular factual settings, especially because of problems associated with measuring the incidence of ‘trade and culture’ conflicts¹⁹ in order to arrive at a meaningful indicator of ‘frequency.’ For example, one relatively straightforward way to measure the frequency of ‘trade and culture’ conflicts would be to look at the number of adjudications in the domain of trade and culture. However, measuring frequency in this manner might be misleading, for as Kaplow (2000, p. 511) pointed out, ‘the frequency of activity may be great even if reported adjudications in a field are few’.

A more promising route for gauging the frequency of ‘trade and culture’ conflicts in specific context consists of probing the specific parameters that are likely to impact frequency itself. In addition to providing an estimate of the likely incidence of ‘trade and culture’ conflicts in a particular context, this approach would facilitate the important task of predicting changes in frequency without having to delve repeatedly into the maze of regulations of one or more jurisdictions. The following represents a tentative list of parameters affecting the frequency of trade and culture conflicts.

- *Market Failures.* Governmental measures aimed at correcting the various types of market failures that arise in the production and distribution of cultural goods and services are likely to lead to commercial tensions to the extent that such measures will often involve recourse to trade- and investment-restricting policies. Generally, the higher the scope for scale economies or the greater the presence of negative externalities affecting domestic cultural industries, the higher the probable frequency of ‘trade and culture’ conflicts.
- *Cultural proximity.* The extent to which foreign cultural products will penetrate a market and exert an effect on domestic culture will partly depend on the degree of cultural proximity between trading partners.²⁰ The concept of ‘cultural proximity’ encapsulates factors such as language, values, beliefs, institutions, and behavioural patterns. Differences in language, for example, will often act as a natural barrier to imports of cultural products, in particular audiovisual products and published materials. So will differences in values, beliefs and behavioural patterns. Such natural barriers will decrease the appeal of imported cultural products and obviate the need to implement additional policies aimed at

¹⁹ In line with this paper’s characterisation of the ‘trade and culture’ problem, a ‘trade and culture’ conflict arises in the presence of a regulation having at least a purported cultural goal that also collaterally restricts trade or investment.

²⁰ For a discussion of a ‘cultural discount’ parameter, see Hoskins, Finn and McFadyen (1997, p. 68).

reducing their availability in the domestic market. Thus, the higher the degree of cultural proximity between jurisdictions, the higher the frequency of ‘trade and culture’ conflicts between them.²¹

- *Trade liberalisation.* The reduction of tariffs and quotas since the inception of the General Agreement on Tariffs and Trade (GATT) in 1947 has been accompanied by a rise in conflicts between trade values and cultural values. This trend can be expected to intensify, especially in light of the firm foundation for future liberalisation opportunities laid by the Uruguay Round negotiations. For example, the ongoing round of talks on services trade and investment liberalisation conducted under the aegis of the GATS and which forms part of the Uruguay Round’s built-in agenda, will likely result in pressures for additional liberalising disciplines that might require policymakers to strike a new balance between trade and cultural values. Generally, then, the increase of trade law (understood here in terms of substantive issue coverage) as well as in the number and quality of scheduled market access and national treatment commitments, could result in a higher frequency of trade and culture conflicts as WTO Members prosecute potential de-liberalising changes in policy through multilateral dispute settlement.
- *Technology.* Recent years have witnessed the rapid transformation of the information and communications technology environment.²² Digital transmission technology drastically expands the capacities to deliver data by many different networks, including terrestrial, cable and satellite. At the same time, the internet is providing a new platform for multimedia services where the traditional systems for the delivery of text, voice, sound and images are merged. Such technological developments are either rendering obsolescent an increasing number of policies aimed at stemming the inflow of cultural products into domestic markets or forcing governments to rethink entirely the role of cultural policies in this new environment. As policymakers seek to respond to the challenges posed by the new information and communications technology environment through the formulation and implementation of new and untested cultural policies, and as the borders between various segments of cultural industries further erode, the frequency of ‘trade and culture’ conflicts is likely to increase.

The parameters discussed above may be applied to particular factual contexts, thus shedding light on the frequency of ‘trade and culture’ conflicts and consequently, on the type of institution — legislative versus adjudicative — to be assigned the task of

²¹ The fractious Canada-US relation in cultural matters offers a case in point.

²² For a more detailed account of the technological impact of globalisation on cultural goods and services, see Footer and Graber (2000).

balancing trade and cultural values. A detailed discussion of where in the WTO institutional structure power over the conflict between trade and cultural values should be exercised is beyond the scope of this paper. Suffice to say that, from a normative standpoint, it is debatable whether the current indeterminacy of WTO law with respect to trade and cultural matters, and the implicit deferral to dispute resolution that it entails (which is itself a reflection of the preference of WTO Members to eschew the difficult bargaining and inevitable ‘concessions’ that a legislated approach would necessitate), is sustainable or indeed desirable. This is especially so at a time when the very legitimacy, both political and juridical, of the WTO is being questioned against the backdrop of domestic anxieties and civil society activism arising from globalisation.

From a positive standpoint, the above analysis suggests that the trade and culture interface can be expected to generate a relatively high incidence of commercial tension, in part because governmental interventions made in response to market failures will often exert restricting effects on imported cultural goods and services given their high tradability. Such a conclusion suggests that future negotiations on cultural goods and services in the WTO could well be characterised by a shift from standards to rules, and that such a shift might be warranted on both efficiency and legitimacy grounds.

In terms of the horizontal allocation of jurisdiction within the WTO institutional structure, such a conclusion implies that the net benefits associated with more specific legislation governing cross-border trade and investment in cultural goods and services are likely to outweigh the political risks and de-legitimising outcomes that primary reliance on the multilateral trading system’s fledgling adjudication mechanism might entail (see, for example, Braun and Parker 1993). In short, the time may be near when members of the international community have to contemplate more seriously the scope for, and the content of, a WTO-anchored set of disciplines aimed at striking an acceptable balance between the potentially competing logics of trade and investment liberalisation on the one hand, and of cultural diversity on the other.

References

- ‘Environnement, Santé, Education, Culture: Naissance des biens publics globaux?’
Courrier de la Planète, vol. 55, pp. 3–4.
- Baker, C.E. 2000, An economic critique of free trade in media products, University of Pennsylvania Law School, unpublished.

-
- Bernier, I. 1998, 'Cultural goods and services in international trade law', in Browne, D. (ed.), *The Culture/Trade Quandry: Canada's Policy Options*, Centre for Trade Policy and Law, Ottawa.
- 2000, *The world trading system and culture: looking at the problem from a cultural perspective*, Faculté de Droit, Université Laval, unpublished.
- and Sauv , P. 1999, *Les industries culturelles et le cycle du mill naire: quels enjeux pour la Francophonie?*, Papier pr par  pour l'Agence de la Francophonie, Paris, unpublished.
- Braun, M. and Parker, L. 1993, 'Trade in culture: consumable product or cherished articulation of a nation's soul?', *Denver Journal of International Law and Policy*, vol. 22, no. 1, pp. 155–71.
- Carmody, C. 1999, 'When 'cultural identity was not an issue': thinking about Canada — certain measures concerning periodicals', *Law and Policy in International Business*, vol. 30, no. 2, pp. 231–320.
- Donaldson, J.D. 1996, 'Television without frontiers: the continuing tension between liberal free trade and European cultural integrity', *Fordham International Law Journal*, vol. 20, no. 1, pp. 90–180.
- Footer, M.E. and Graber, C.B. 2000, 'Trade liberalisation and cultural policy', *Journal of International Economic Law*, vol. 3, no. 1, pp. 47–81.
- Hoskins, C., Finn, A. and McFadyen, S. 1997, 'Television and film in a freer international trade environment: US dominance and Canadian responses', in McAnany, E.G. and Wilkinson, K.T. (eds), *Mass Media and Free Trade: NAFTA and the Cultural Industries*, University of Texas Press, Austin, pp. 63–91.
- Kaplow, L. 1992, 'Rules versus standards: an economic analysis', *Duke Law Journal*, vol. 42, no. 3, pp. 557–629.
- 2000, 'General characteristics of rules', in Bouckaert, B. and De Geest, G. (eds), *Encyclopedia of Law and Economics*, Edward Elgar Publishers, Cheltenham.
- Krugman, P. and Obstfeld, M. 1997, *International Economics: Theory and Policy*, 4th edn, Addison-Wesley, Reading, Massachusetts.
- Nicolaidis, K. and Trachtman, J.P. 2000, 'From policed regulation to managed recognition in GATS', in Sauv , P. and Stern, R.M. (eds), *GATS 2000: New Directions in Services Trade Liberalization*, Brookings Institution and Center for Business and Government, Harvard University, Washington DC.
- Olson, M. 1965, *The Logic of Collective Action: Public Goods and the Theory of Groups*, Harvard University Press, Cambridge, Massachusetts.

-
- Rony, H. 2000, 'Quotas: effet dopant?', *Courrier de la Planète*, vol. 55, no. 2, pp. 52–3.
- Schwanen, D. 1998, 'The summer of our (Canadian) content: thinking about Canada's response to the WTO magazine decision', *Backgrounder*, C.D. Howe Institute, Toronto.
- Shao, W.M. 1995, 'Is there no business like show business? free trade and cultural protectionism', *Yale Journal of International Law*, vol. 20, no. 1, pp. 105–150.
- Sykes, A.O. 2000, Efficient protection through WTO rulemaking, Paper presented at 'Efficiency, Equity and Legitimacy: The Multilateral Trading System at the Millennium', John F. Kennedy School of Government, Harvard University, 1–2 June.
- Trachtman, J.P. 1998, 'Trade and ... problems, cost-benefit analysis and subsidiarity', *European Journal of International Law*, vol. 9, no. 1, pp. 32–85.
- 1999, 'The domain of WTO dispute resolution', *Harvard International Law Journal*, vol. 40, no. 2, pp. 333–78.
- Wernick, M. 2000, Global diversity: myths and global challenges, Paper presented at a conference on 'Prospects for Culture in a World of Trade', New York University, 3 March.
- WTO (World Trade Organization) 1998, 'Audiovisual services, Background note by the Secretariat', Sectoral Paper S/C/W/40, Council for Trade in Services, WTO, Geneva.

Discussant — *Patrick Messerlin*

Institut d'Etudes Politiques de Paris

GATS negotiations on audiovisual commitments

The paper by Pierre Sauvé mentioned that only 18 countries had scheduled commitments for audiovisual services in the Uruguay Agreement. But the countries that have signed the agreement include the United States and India, between them the largest producers of movies, as well as Japan, Mexico and Hong Kong. The only large producer of movies that has not signed the agreement is Brazil, though it could well join in the future. English and Spanish are represented in the agreement and Portuguese may soon be introduced if Brazil becomes a signatory. In a sense, the European Community is already present on the list of signatories in terms of language. The language links may encourage countries like the United Kingdom, Portugal and Spain to join.

Culture and entertainment

Culture is not entertainment. It is relatively easy to draw a line between culture and entertainment. It is uncertain what will be the culture of the early 21st century — people will only know this in 50 years time when they look at the works that have survived and acknowledge their worth. The lesson of history is that few people have been able to recognise the masters of their time in terms of culture.

Economic features

The paper mentioned the economic features of audiovisuals — public goods, scale economies, externalities and collective action.

But movies are not public goods in a theatre. Individuals pay for a seat.

There are no scale economies in terms of production. Indeed, the US major companies are producing fewer films than are the three French majors (the high number of French films is also related to the subsidies received by French filmmakers). It is the actors rather than the directors that are the integral part of movie production today, and it is difficult to talk about scale economies in terms of actors. Pierre Sauvé is correct in saying that there are scale economies in distribution. But in most European countries, American firms do not own the distribution network. Rather, it is owned jointly by the American and European firm, so the European firms are part of the story.

Even if scale economies were large, the development of new technology such as the Internet means that there are no longer barriers to entry in distribution. If there is one country in the world that should not put up barriers to the internet, it is France — the internet enables French movies to reach French-speaking people in Okalahoma, New South Wales or Africa at very low cost. As technology progresses, the US distribution system is loosing the advantage of size, and this creates a good opportunity to compete in the market.

It is possible that there are externalities associated with audiovisuals, but are these positive or negative and what instruments could be used to manage these externalities? It is too complex to handle and it is doubtful whether there are any externalities at all.

One point that is more important than public goods, scale economies or externalities is product differentiation. The paper talks about one unique industry, but Hollywood and the production of cultural films are clearly different industries. The maker of a cultural film does not dream of 10, 15 or 20 million viewers during the six months following the film's release, but, rather, dreams of having 10 or 20 million viewers over the next 50 years. The filmmaker knows they are making a different film that may not attract a large audience.

Because of product differentiation, the Europeans are becoming exporters of audiovisual works such as game shows and sitcoms, not just movies. It is not yet a two-way trade, but there is trade in differentiated products.

The last 40 years suggests that protection has Americanised the French quicker than free trade. Protection attracts resources into areas in which others have a comparative advantage, and this is the last thing you should want to do when you are proud of your culture.

General discussion

The discussion on the papers presented in chapters 12 and 13 focused on:

- defining culture;
- the economics of protecting culture;
- patronage as a domestic policy to protect culture;
- subsidies to protect culture;
- the review of the Australian Broadcasting Services Act; and
- developing a Reference Paper on audiovisual services.

Defining culture

It was noted that the Australian regulatory system did differentiate between cultural and industrial material. There was a sliding scale, whereby to qualify as Australian content, material had to meet certain cultural criteria about such things as the creative principles involved. Co-production activity attracted different levels of points. Finally, there was a range of audiovisual production in Australia that did not receive, nor seek, subsidies or regulatory protection of any sort. This last area of Australian production was growing in a significant way. So there were already practical ways of differentiating industrial from cultural activity.

The economics of protecting culture

It was noted that there had been a study of the costs and benefits of Australian content in television that had taken externalities into account, and found that the costs were commensurate with the benefits (Papandrea 1997).

It was suggested that Pierre Sauvé's economic case for protecting culture ignored two significant factors.

First, monopolistic competition with endogenous product differentiation would typically lead to excessive product variety, rather than too little product variety. And with monopolistic competition, the costs of measures such as quotas would be especially high.

Second, and more importantly, intellectual property rights already dealt with many of the market failures identified with cultural works. In general, intellectual property protection was sensible, but if the system had an element of precedence, that again would lead to excessive product variety.

If the cultural goods were non-rivalrous in consumption, then duplication of those goods as a result of quotas would amount to pure waste. If, in addition to being non-rivalrous, there were externalities associated with incorrect production of the good, then the combination of the pure productive waste from duplication and the externalities associated with producing the wrong goods would make the costs of remedies very high relative to the problems being addressed.

It was noted that both papers failed to address the fact that through copyright laws, virtually every country with the exception of New Zealand restricted the parallel importation of copyrighted goods. These restrictions were a significant source of indirect subsidy to the parties that controlled the goods. With monopolistic competition, these restrictions induced allocative and technical efficiency losses.

Patronage as a domestic policy to protect culture

A participant advocated domestic patronage as a means of protecting culture, suggesting the relevant market failure was potential myopia in private tastes. The present value of Van Gogh's paintings was huge, yet nobody would have lent Van Gogh money in his own lifetime. Indian classical music would not have survived either without patronage.

If there was a case for patronage, how was it best implemented? Patrick Messerlin thought the British way of funding culture was interesting, whereby money from the National Lottery was used to subsidise cultural activities. An alternative option to fund domestic film production would be by taxing moviegoers. Boards, like film festival boards, could decide which filmmakers had 'potential' and should therefore receive subsidies. Patrick Messerlin reiterated that he did not believe democratic governments could handle the support of culture, and referred to the successes of the British non-government patronage system involving competition between patrons.

Subsidies to protect culture

Patrick Messerlin noted that although quantitative restrictions had a positive impact on the music industry currently, this was unlikely to continue. Eventually, subsidies would be captured and unsuccessful artists would remain in the industry just to receive subsidies.

Patrick Messerlin also suggested that French subsidies were of concern, not because of the GATS or criticisms from the United States, but because some people thought

that this was the only way to create French moviemakers. He was concerned about subsidies for cultural impact reasons, not for trade negotiation reasons.

Review of the Australian Broadcasting Act

The cultural objectives of broadcasting regulation were a major part of the terms of reference for the Productivity Commission's inquiry into the Broadcasting Services Act and Australian broadcasting policy (PC 2000). The pursuit of cultural objectives was alleged to have a pervasive effect on broadcasting policy. This was evident through a series of quid pro quos relating to restrictions on entry into television and radio, and to the obligations placed on broadcasters to incur extra costs in adhering to a quota system for local programming (despite the fact that the costs of Australian productions involved was minute compared with the value of rents obtained from the restrictions on entry into the industry).

The Productivity Commission's report looked at the effects of barriers to entry and the cultural policy issues associated with content quotas. The barriers to entry into broadcasting had created monopolistic conditions within audiovisual markets, with the perverse effect of actually reducing the demand for local culturally significant programming. Producers of audiovisual material claimed that the small number of potential buyers for their product meant that they were being constantly squeezed on prices and were increasingly forced to reduce production values. This, in turn, reduced the attractiveness of the products they were able to make relative to competing imports. So there was a perverse relationship, rather than the quid pro quo claimed in Australian policy.

The Productivity Commission also looked at the issue of quotas and came to similar conclusions as Pierre Sauvé. The rapid changes in technology and the diversification of platforms for the dissemination of audiovisual material were likely to have profound effects on the industry and the ability of past measures to be successful in achieving cultural policy objectives. The report recommended there be a review of policy measures for this purpose by 2004.

Reference Paper on audiovisual services

One participant doubted whether there were sufficient countries in support of a Reference Paper on audiovisuals. Of the 19 Members that had made WTO commitments, some were major film producers. But only two countries had made full commitments — the United States and the Central African Republic. Most major film producers had not made anything close to full commitments on audiovisual services.

Australia had taken cultural objectives into account, and so had not made commitments on audiovisual services and had taken out most favoured nation (MFN) exemptions for bilateral co-production agreements. The MFN exemptions were subject to a theoretical ten-year sunset rule that expired in 2005. But there was little support to remove these exemptions and the process of negotiating them away was likely to be slow. It was also observed that the GATS did not currently have disciplines on either subsidies or domestic regulation. It would be a long time before there was sufficient momentum to create a Reference Paper on audiovisual services addressing these issues.

Pierre Sauvé supported this pessimistic view on both subsidies and the MFN derogation, noting the recent signals from Geneva and the repeated failures of the multilateral community to tackle subsidy issues.

References

- Papandrea, F. 1997, *Cultural Regulation of Australian Television Programs*, Bureau of Transport and Communications Economics Occasional Paper no. 114, AGPS, Canberra.
- PC (Productivity Commission) 2000, *Broadcasting*, Report no. 11, AusInfo, Canberra.

PART F

MEASURING AND
MODELLING
THE IMPACT OF
RESTRICTIONS ON TRADE IN
SERVICES

14 ‘Sins of commission and omission’: measuring regulatory impediments to trade in services

Tony Warren and Christopher Findlay[†]

Network Economics Consulting Group Pty Ltd and Australian National University

14.1 Introduction

Transparency of policy affecting services trade and investment is critical for successful reform. More information, in terms of the detail of policy and the analysis of its effects, helps mobilise the key countervailing interests against the protectionist forces in domestic economies, facilitates the construction of coalitions for reform by political leaderships, and adds to policymaker confidence as strategies are designed and implemented. Yet there is little of this type of work available for the service sector.

This paper reviews the work undertaken to date on identifying impediments to trade and investment in services and suggests where this research can be extended.

Specifically, this paper reviews work detailed in Findlay and Warren (2000) that identifies the impact on trade and investment in services of impediments that can generally be grouped under the rubric of ‘entry restrictions’. These are primarily policy limitations on the number of firms that can operate within an industry, or restrictions on the level of foreign investment allowed in incumbent firms.

These restrictions have a significant and ongoing impact on economic outcomes (prices, costs and/or quantities) in countries around the globe.

During the course of the research into the impact of entry restrictions, it became apparent that a significant number of (mainly developed) countries had removed most of these restrictions on entry. Yet anecdotal evidence suggests that impediments to trade and investment remain in these countries. Specifically,

[†] This paper draws heavily on Warren and Findlay (2000).

potential entrants routinely complain about a lack of *appropriate* regulation designed to ameliorate such ‘market failures’ as asymmetrical information, the need for standardisation and the abuse of market power.¹ Whereas most research has focused on the measurement of entry barriers — the ‘sins of commission’ — this paper also canvasses what would be required to quantify the impact of a failure to regulate — ‘sins of omission’.

14.2 What are barriers to trade in services?

Services transactions

A service is *an economic activity that adds value, either directly to another economic unit or to a good belonging to another economic unit.*² Consequently, a defining feature of services is the requirement for direct interaction between producers and consumers (or at least a consumer’s assets) before the service can be rendered (Hirsch 1989).

The need for producers and consumers to interact to render a service influences how international transactions in services are conducted. If a preferred service producer is located in another country, then a consumer must somehow interact with that producer to acquire those services. Article I (scope and definition) of the General Agreement on Trade in Services (GATS), substantially following Bhagwati (1984) and Sampson and Snape (1985), developed a four-part typology of how such capabilities can be accessed internationally:

- through cross-border communications in which neither the producer nor the consumer moves physically, interacting instead through a postal or a telecommunications network;
- through the movement of a consumer to a supplier’s country of residence;
- through the movement of a commercial organisation to the consumer’s country of residence; or
- through the movement of an individual service supplier to the consumer’s country of residence.

Consequently, the concept of international service transactions encompasses foreign direct investment and the movement of labour, as well as traditional cross-border

¹ Allegations are also commonly made that limited institutional capability or outright corruption undermines whatever formal policy commitments are made. These issues are not addressed in this paper.

² This definition is derived from the classic definition of services proposed by Hill (1977, p. 317).

transactions. In this paper, any policy that impedes service producers and consumers interacting through any of these channels (or modes of supply) is considered an impediment to international service transactions.

Categories of impediments

It is common practice to categorise impediments to trade in services into market access restrictions and derogations (departures) from national treatment.³ Part III of the GATS explicitly introduces the concepts of *market access* and *national treatment* into the international services trade architecture. Surprisingly, the GATS does not specifically define market access. Article XVI.1 simply obliges members to grant market access to scheduled industry sub-sectors, while Article XVI.2 contains a list of measures considered to be limitations on market access. Article XVII.1 defines national treatment as treatment no less favourable than that accorded to like domestic services and service providers, subject to the limitations and conditions set out in the country's schedule of commitments.

The implication of Part III is that market access and national treatment are broader in scope than the corresponding market access and national treatment provisions in the GATT (Hoekman 1995 and Snape 1998). To begin with, the GATS provision on national treatment does not draw a distinction between frontier and internal constraints, but embraces all policies that might discriminate between domestic and foreign suppliers. In contrast, national treatment in the GATT extends to matters of internal taxation and regulation only. In effect, the GATS article on national treatment encompasses both national treatment and market access as normally defined.

More importantly, the GATS article on market access extends beyond traditional concerns of access for foreign service suppliers to encompass all policies that restrict access to a market. This is a major extension of multilateral trade disciplines into the realm of domestic policy, in particular competition policy. As Snape (1998) suggests, the 'GATT is almost entirely concerned with relations between "us" and "them"; these provisions of [the] GATS are not concerned with "us" and "them" but [with] "some of us" on the one hand and "the rest of us and them" on the other'.

The lack of clarity within the GATS as to what is meant by these important concepts has meant that the process of categorising specific policies into limitations on market access or breaches of national treatment has proven to be particularly difficult. Following the insights of Snape (1998), Findlay and Warren (2000)

³ UNCTAD and World Bank (1994, chapters 4–7) provide an early example of such distinctions.

propose to operationalise the distinction between market access and national treatment by focusing on the concept of discrimination.

- *Market access* means non-discrimination between incumbents in a particular market and possible entrants (be they domestic or foreign). Hence, a legislated monopoly is considered a market access limitation.
- *National treatment* means non-discrimination between domestic and foreign service suppliers. Therefore, a policy limiting foreign investment is considered a breach of national treatment.

Failing to regulate as a barrier to trade

This definition of what constitutes a barrier to trade, while representing a radical extension of the traditional concept of border restrictions, still refers to active policies on the part of governments to favour incumbents over entrants, or domestic producers over foreign producers. Increasingly, there is scope for extending the concept of trade barriers still further to include the failure of governments to act to overcome market failures that have the effect of discriminating against potential entrants, be they foreign or domestic.

In many service industries, the lack of appropriate essential facilities regulation is commonly cited as the most significant sin of omission. Essential facilities regulation refers to a set of legislative instruments that provide open access to an essential input into the production process, owned and operated by another firm, that exhibits natural monopoly characteristics.⁴ This form of regulation first arose with the common law doctrine of *prime necessities*, which is generally traced back to a mid-17th century case that established the principle in the case of monopoly wharfs. Lord Hale made the point that if people were constrained to use a particular private wharf because it was the only wharf at the port of the licensed wharf, then the wharf owner had an obligation not to set arbitrary and excessive or immoderate duties. That is, as an obligation under common law, monopolies providing essential commodities must charge reasonable and moderate prices. In Australia, this doctrine has been enshrined in legislation via Parts IIIA and XIC of the *Trade Practices Act 1974* (Cwlth).

Essential facilities legislation is of critical importance in relation to service industries. Indeed, the operation of the essential facilities components of the

⁴ A natural monopoly occurs when one enterprise can supply a service or a bundle of services at a lower cost than two or more enterprises, generally as a consequence of economies of scale or scope (see Baumol, Panzar and Willig 1982). A monopoly may also occur as a result of significant barriers to entry (high start-up costs) and barriers to exit (high costs of converting sunk assets to alternative uses).

Australian legislation is explicitly limited to services. Most services that require access to a network, such as telecommunications, gas, electricity, water and transport, find that some part of that network exhibits natural monopoly characteristics (at least in the short to medium term). In countries where effective essential facilities legislation is not operating, potential entrants into these industries often claim that they cannot compete effectively with incumbent operators that control the natural monopoly components of the network. For these competitors, such sins of omission can operate to prevent entry into a market in much the same way as constraints on licensing.

14.3 Measuring sins of commission

Since the conclusion of the Uruguay Round, significant steps have been made in identifying and quantifying the impediments to trade and investment in services that can generally be grouped under the rubric of ‘entry restrictions’. Following the method developed as part of an Australian research project on this issue (Findlay and Warren 2000), the measurement of impediments to trade and investment in services involves a three-step process.

- First, available qualitative evidence is collected on the way nations discriminate against potential entrants in various service industries. This evidence is transformed into a frequency-type index, with every attempt made to weight discriminatory policies by their economic significance.
- Second, the impact of the policies, as measured by the frequency indexes, is assessed against cross-national differences in domestic prices, costs and/or quantities, with the effect of other factors explaining cross-national differences explicitly taken into account.
- Third, the measured impact of the frequency indexes on prices, costs or quantities is incorporated into a general equilibrium model to assess the economywide impacts of the policies at issue. Where possible, partial equilibrium modelling is also undertaken to allow for the ‘black box’ liberalisation to be more clearly understood.

Frequency indexes

Frequency indexes have been used extensively for the quantification of non-tariff barriers (NTBs) affecting trade in goods. Frequency-type measures of NTBs affecting services, however, were not produced for many years because of the lack of suitable data on impediments. This lack of data was partially overcome thanks to the requirement of the GATS that countries agreeing to be bound by the multilateral

trade disciplines would list in their individual GATS schedules those sectors in which they were prepared to make commitments, and any specific barriers they wished to retain.

Quantification of the GATS schedules commenced with the pioneering work of Hoekman (1995), who developed a three-category weighting method. He examined all GATS schedules and, for quantification purposes, allocated a number to each possible schedule entry (ie each possible market access or national treatment commitment in each mode of delivery in each industry sub-sector). Hoekman used these measures as a mechanism for quantifying the extent of commitments (the greater the number, the more commitments made). Other researchers have inverted the analysis and examined the number of commitments that have not been made (the greater the number, the more illiberal the economy).

There are many limitations with the Hoekman methodology. To begin with, it does not distinguish between barriers in terms of their impact on the economy — minor impediments receive the same weighting as an almost complete refusal of access. Moreover, there are severe limitations with the coverage of the GATS schedules, which in many cases do not give an accurate picture of the actual barriers in place. This is particularly the case for developing economies, many of which were unable to provide the detail required to meet the complexities of the scheduling process. There is also some evidence to suggest that nations with liberal policies left some services unbound so as to maintain a retaliatory capability in future market access negotiations. Therefore, some industries that are recorded in the Hoekman indices as being impeded may be open, at least to suppliers from some economies.

Various studies, many produced as part of the Australian services project, have attempted to develop, at a sectoral and modal level, a more complex weighting system than that used by Hoekman, seeking to quantify differences in the effect of different partial commitments. More extensive databases have also been drawn upon to overcome some of the informational limitations with the GATS schedules.

Hardin and Holmes (1997), for example, include in their analysis on the types of barriers affecting investment the information contained in the Individual Action Plans (IAPs) produced by members of the Asia Pacific Economic Cooperation (APEC) process. These documents do provide additional information but are still far from an exhaustive description of impediments. They also sought to incorporate the relative economic impact of different policies into frequency data on the types of barriers affecting investment. Five types of barriers to foreign investment were identified. Weights were then developed within each of those categories on the basis of the perceived economic impact of each policy category. For example, they give a much greater weight to a policy that completely excludes foreign equity than to a policy that allows more than 50 per cent but less than 100 per cent foreign equity.

Policies that limit investment in existing firms, but allow greenfields investment are given a lower weight than policies that limit all investment.

McGuire (1998) graphically demonstrates the limitations of using information derived from international agreements, such as the GATS and the APEC Individual Action Plans. McGuire undertook a detailed analysis of the Australian financial service policy regime (at both State and Federal level) in an effort to get a clearer picture of the measures affecting financial services trade. He includes, where applicable, prudential regulations within the definition of market access and national treatment restrictions. McGuire uncovers a host of measures, including government monopolies over the provision of certain types of financial services, prudential regulation, restrictions on direct foreign investment in banking and insurance, discriminatory government licensing requirements and government guarantees to selected financial service providers. In total, some 165 measures are identified, compared with the 38 financial service impediments listed in Australia's GATS schedule. McGuire then applies the weighting methodology developed by Claessens and Glaessner (1998) in their study of GATS impediments to financial services to his list of impediments for Australia.

In relation to telecommunications, Warren (2000a) used a 1997 survey by the International Telecommunications Union (ITU) to construct a set of policy indices for 136 countries. These data have the advantage of being drawn from a survey of actual policies, rather than inferring these policies from commitments made in trade negotiations. Five separate indices were constructed, corresponding with the more important distinctions drawn in the GATS context, namely the differences between market access and national treatment and between trade and investment. In the construction of these indices, Warren sought to incorporate economic as well as legal inputs by including a count of the number of firms actually competing in a market, as well as measures of the formal policies.

In the transport sector, McGuire, Schuele and Smith (2000) developed a technique for assessing impediments to trade in maritime services. The data on policy came from a variety of sources including a questionnaire developed by the WTO's Negotiating Group on Maritime Services, GATS schedules, WTO Trade Policy Reviews, information from the office of the US Trade Representative, material published by the OECD and APEC IAPs. Different types of impediments were classified and weights were given to them on the basis of the perceived economic impact of each policy category. Nguyen-Hong (2000) applied a similar methodology to professional services — specifically legal, accounting, architectural and engineering services — and Kalirajan (2000) to distribution services.

Partial impact measures

Having identified and systematised the various cross-national differences in policy, it is possible in some industries to estimate the impact of these differences on core economic outcomes, such as prices and consumption. The available research on the measurement of NTBs affecting goods trade provides a useful starting point. In their extensive review of the literature, Deardorff and Stern (1985) identify two broad methods of quantifying the economic impact of NTBs.

- *Price impact measures* that examine the impact of NTBs on domestic prices by comparing them with world prices.
- *Quantity impact measures* that compare an estimate of trade volumes in the absence of NTBs with actual trade volumes.

These types of price and quantity impact measures have been considered impossible to replicate in relation to service industries on grounds of data availability. A world price for many service industries is indeterminate. Similarly, lack of systematic bilateral services trade data and the highly aggregated nature of current account data limit the potential for traditional quantity impact models (Ascher and Whichard 1991).

As a consequence of these data concerns, it is necessary to identify alternative benchmarks against which to compare actual prices and quantities. Here the market power analysis associated with competition or anti-trust regulation is instructive. The aim of such analysis is to compare actual market outcomes with those that would be expected to prevail if the market were competitive (Areeda, Hovenkamp and Solow 1995).

Price impact measures compare the observed prices in an economy with those likely to prevail in the absence of impediments to trade and investment. The theory underlying the approach is that if the market had no impediments to entry, then it would be competitive and prices would be expected to approach a firm's long-run marginal cost — defined as the cost of keeping a particular facility alive and well in the long run.⁵ If there are impediments, however, there will be a wedge between price and marginal cost. Impediments can affect not only the size of the margin of price over costs, but also costs themselves. Costs might be higher because low cost suppliers are excluded from the market or because protected firms are not operating at their lowest possible cost levels. For all these reasons — the margin effect, the cost difference effect and the cost reduction effect — prices observed in the presence of impediments may exceed those in their absence.

⁵ The excess of price over marginal cost as a proportion of price is known as the Lerner Index (Lerner 1934).

As a first step in estimating the impact on prices of impediments to trade in services, Kalirajan et al. (2000) examined price-cost margins (or ‘net interest margins’) of 694 national and State commercial banks in 27 economies. Using a two-stage econometric technique, they were able to isolate the specific impact on this margin of the trade restrictiveness indices developed by McGuire and Schuele (2000), while correcting for the factors that influence the size of the buffer that banks need to manage their risks. Kalirajan (2000) and Nguyen-Hong (2000) have completed similar studies for distribution and engineering services.

An alternative approach to focusing on prices is to examine output, since price and output is simultaneously determined in a market. In particular, demand for various services is likely to be greater, the more competitive is its supply. This is because the lower relative prices and higher service quality arising from competition will increase demand, while rivalry in investment will push out supply.

Warren (2000b), for example, quantified the comparative impact on telecommunications consumption of limits on competition, controlling for other explanatory variables. Consumption is measured in terms of the number fixed and mobile connections per hundred persons. Restrictions on competition are modelled using the ITU-derived indices of telecommunication policies discussed above.

Modelling

There are three main limitations with the price and quantity approaches to assessing the economic impact of impediments to trade and investment in services.

- First, the nature of the predicted reductions in price or increases in quantity remains a black box. Does liberalisation affect economic outcomes through competition effects (reduced market power), cost reduction effects (increased efficiency) or cost differential effects (reduced input prices)?
- Second, such measures fail to give a picture of the welfare implications from reform, including impacts on consumer and producer welfare.
- Third, these techniques fail to capture important intersectoral effects. What impact does liberalisation of a particular service industry have on the economy as a whole?

To overcome these limitations it is necessary to use these partial equilibrium observations to inform a general equilibrium modelling exercise. The accompanying paper by Dee, Hanslow and Phamduc discusses this step in the analysis at length (chapter 15 in this conference volume).

Summary

Significant advances have been made in recent years in the identification and quantification of the set of policies restricting market entry. Unfortunately, there has been substantially less progress made in identifying how the absence of particular policies impedes market access. In part, this is a function of the relative importance of different impediments. A lack of essential facilities legislation is irrelevant if the Government confines access to a market to just one service provider. But there is also the analytical problem of identifying what is and what is not effective regulation.

14.4 Measuring sins of omission

The nature of effective regulation

Across the whole spectrum of economic regulation, there is very little consensus on what effective regulation actually looks like. There is significant disagreement on what legislative form and set of procedures will produce optimal economic outcomes. This is particularly so for essential facilities regulation, with the overwhelming majority of countries around the world not even feeling the need for such regulation.

Theoretically, optimal economic regulation is a policy where the incentives are structured such that the sum of administrative and error costs are minimised (Cooter and Ulen 1997; Landrigan and Warren 2000).

- *Administrative costs* refer to the costs incurred by parties to an action in prosecuting and defending a case or in pursuing and influencing regulatory intervention.
- *Error costs* are the efficiency costs to society of incorrect decisions on whether or not to intervene in the market. Most regulators have a discretionary power over whether to intervene. In making this decision, the regulator must weigh the social costs of falsely condemning competitive behaviour or identifying market failure (Type I error) against the costs of erroneously exonerating anticompetitive conduct or failing to correct market failure (Type II error).

If policymakers believe that the scope for regulatory failure is high and the costs of market failure are low, then the sum of administrative and error costs is minimised through a light-handed regulatory regime. Conversely, if the costs of market failure are very high and the scope for regulatory failure minimal, then the sum of

administrative and error costs is minimised through a more interventionist regulatory regime.

In practice, it is very difficult to assess *a priori* the cost of market failure and the scope for regulatory failure. Moreover, it is very likely that the costs and benefits of intervention will differ dramatically across time and circumstance. This has significant implications for the purpose of measuring the impact of a failure to regulate on trade and investment in services. Unlike entry restrictions, where their removal leads to better economic outcomes, there is no simple linear relationship between the amount of regulation and economic outcomes over the longer term. A more interventionist policy does not necessarily lead to better economic outcomes.

A methodology for wider analysis

To assess the point at which regulation is effective would appear to require, at a minimum, the adoption of a structure-conduct-performance methodology and its application across a range of countries. The structure-conduct-performance approach requires:

- delineation of the key *structural elements* considered most likely to drive the performance of the regime;
- identification of the *incentives* that alternative elements create for the major actors involved with the regime; and
- assessment of the cross-national impact of these incentives on key industry *outcomes*.

Identification of the key structural elements of a regulatory regime is perhaps the most difficult part of the structure-conduct-performance methodology, both in theoretical and empirical terms. However, drawing upon the theory of what constitutes optimal economic regulation (discussed above), it is possible to identify a tractable set of parameters. Essentially, researchers need to identify the regulatory elements that maximise the scope for market failure to go unpunished (Type II errors) versus those elements that maximise the scope for regulatory failure (Type I errors).

To identify the scope for market failure to go unpunished, the first step is to specify the market failure and the requisite regulatory response. This step is far from uncontroversial, with analysts around the world seldom agreeing on the existence of market failure, let alone the optimal regulatory solution. Nevertheless, it is possible in some clearly defined cases, such as access to essential facilities, to obtain some consensus on the scope for market failure and the need for some form of regulated access. In countries where no such regulation operates, it is safe to assume that in

situations where the incumbent controls monopoly elements of the core network, the scope for market failure (ie incumbents taking advantage of the significant power afforded by their control of those essential facilities) is relatively high. More generally, if the hurdles to regulatory intervention are high then the scope for Type II errors increases.

To identify the scope for regulatory failure (Type I errors), it is important to focus on the independence of the regulator. Arguably, the greater the discretion afforded the regulator and the greater its powers, the more likely it is that regulatory failure will be observed (Bhagwat 1999). Measuring this discretion is a difficult task, but some indication may be obtained by seeking information on the discretion the regulator has in terms of the scope of the regime and its ability to set terms and conditions. More generally, if the hurdles to regulatory intervention are low then the scope for Type I errors increases.

Once the key structural elements of the regulatory regime are delineated, the impact of the chosen structural variables on the incentives facing the regulator and industry participants needs to be explicitly identified. In relation to essential facilities regulation, the incentives created by differing regulatory structures are relatively easy to identify. An access regime that is particularly onerous on the access provider will facilitate, in the short term, competitive entry using incumbent facilities, with consequent impacts on prices, quantities and market shares. In the longer term, however, efficient investment in infrastructure will be discouraged and the access seeker will have reduced incentives to upgrade existing facilities.⁶ An access regime that is more tightly constrained is likely to have the opposite impact on incentives.

Thus, it is apparent that the set of performance indicators necessary to assess the effectiveness of a regulatory regime must extend beyond price and quantity data to include such variables as investment and market share. This adds significantly to the data collection task, but it would seem that any analysis focusing solely on prices and quantities would be at risk of failing to identify the costs of regulatory failure.

The joint research effort on services undertaken by the Australian National University and the Productivity Commission is in the early stages of extending this research to a broader set of service industries. It is hoped that the application of such a methodology on a large enough sample of economies will allow for a

⁶ The impact of any increased risk of Type I errors on investment can be seen in the analysis of Bittlingmayer (1999) who examined a set of 21 major industries in the United States from 1947–91 to investigate the statistical association between antitrust case filings and investment. Each extra antitrust case filing was found to be associated with a significant decline in investment in the industry at issue. There are many potential causes for this association, but the increased risk associated with antitrust activity, including the risk of Type I errors, is considered pivotal.

rigorous assessment of the extent to which various regulatory structures affect services trade and investment. In combination with the research already completed, it is hoped that this research will lead to a fuller assessment of the impediments affecting services trade and investment — both the sins of commission and omission.

References

- Areeda, P., Hovenkamp, H. and Solow, J. 1995, *Antitrust Law: An Analysis of Antitrust Principles and their Application*, Little, Brown and Company, Boston.
- Ascher, B. and Whichard, O. 1991, 'Developing a data system for international sales of services: progress, problems and prospects', in Hooper, P. and Richardson, J. (eds), *International Economic Transactions: Issues in Measurement and Empirical Research*, University of Chicago Press, London.
- Baumol, W., Panzar, J. and Willig, R. 1982, *Contestable Markets and the Theory of Industry Structure*, Harcourt, Brace, Jovanovich, New York.
- Bhagwat, A. 1999, 'Modes of regulatory enforcement and the problem of administrative discretion', *Hastings Law Journal*, vol. 50, April, pp. 1–60.
- Bhagwati, J. 1984, 'Splintering and disembodiment of services and developing countries', *World Economy*, vol. 7, no. 2, pp. 133–44.
- Bittlingmayer, G. 1999, 'Investment and antitrust enforcement', <http://www.gsm.ucdavis.edu/~gnbittli/>
- Claessens, S. and Glaessner, T. 1998, *Internationalization of Financial Services in East Asia*, World Bank, Washington DC.
- Cooter, R. and Ulen, T. 1997, *Law and Economics*, 2 edn, Addison-Wesley, Reading, Massachusetts.
- Deardorff, A. and Stern, R. 1985, *Methods of Measurement of Non-Tariff Barriers*, United Nations Conference on Trade and Development, Geneva.
- Findlay, C. and Warren, T. (eds) 2000, *Impediments to Trade in Services: Measurement and Policy Implications*, Routledge, London and New York (forthcoming).
- Hardin, A. and Holmes, L. 1997, *Services Trade and Direct Foreign Investment*, Industry Commission Staff Research Paper, AGPS, Canberra.
- Hill, T.P. 1977, 'On goods and services', *Review of Income and Wealth*, vol. 24, no. 4, pp. 315–38.
- Hirsch, S. 1989, 'Services and service intensity in international trade', *Weltwirtschaftliches Archiv*, vol. 125, pp. 45–60.

-
- Hoekman, B. 1995, 'Tentative first steps: an assessment of the Uruguay Round agreement on services', Policy Research Working Paper no. 1455, World Bank, Washington DC.
- Kalirajan, K. 2000, *Restrictions on Trade in Distribution Services*, Productivity Commission Staff Research Paper, AusInfo, Canberra.
- , McGuire, G., Nguyen-Hong, D. and Schuele, M. 2000, 'The price impact of restrictions on banking services', Chapter 13 in Findlay, C. and Warren, T. (eds), *Impediments to Trade in Services: Measurement and Policy Implications*, Routledge, London and New York (forthcoming).
- Landrigan, M. and Warren, T. 2000, 'Administrative costs and error costs in market conduct regulation: two case studies', *Competition and Consumer Law Journal*, vol. 7, no. 3, pp. 224–39.
- Lerner, A.P. 1934, 'The concept of monopoly and the measurement of monopoly power', *Review of Economic Studies*, vol. 1, pp. 157-188.
- McGuire, G. 1998, *Australia's Restrictions on Trade in Financial Services*, Productivity Commission Staff Research Paper, AusInfo, Canberra.
- and Schuele, M. 2000, 'Restrictiveness of international trade in banking services', Chapter 12 in Findlay, C. and Warren, T. (eds), *Impediments to Trade in Services: Measurement and Policy Implications*, Routledge, London and New York (forthcoming).
- , —— and Smith, T. 2000, 'Restrictiveness of international trade in maritime services', Chapter 10 in Findlay, C. and Warren, T. (eds), *Impediments to Trade in Services: Measurement and Policy Implications*, Routledge, London and New York (forthcoming).
- Nguyen-Hong, D. 2000, *Restrictions on Trade in Professional Services*, Productivity Commission Staff Research Paper, AusInfo, Canberra.
- Sampson, G. and Snape, R. 1985, 'Identifying issues in trade in services', *World Economy*, vol. 8, no. 2, pp. 171–82.
- Snape, R. 1998, 'Reaching effective agreements covering services', in Krueger, A. (ed.), *The WTO as an International Organization*, University of Chicago Press, Chicago and London.
- UNCTAD (United Nations Conference on Trade and Development) and the World Bank 1994, *Liberalizing International Transactions in Services: A Handbook*, United Nations, New York.
- Warren, T. 2000a, 'The identification of impediments to trade and investment in telecommunication services', in Findlay, C. and Warren, T. (eds), *Impediments*

to Trade in Services: Measurement and Policy Implications, Routledge, London and New York (forthcoming).

—— 2000b, ‘The impact on output of impediments to trade and investment in telecommunication services’, in Findlay, C. and Warren, T. (eds), *Impediments to Trade in Services: Measurement and Policy Implications*, Routledge, London and New York (forthcoming).

—— and Findlay, C. 2000, ‘Measuring impediments to trade in services’, in Sauvé, P. and Stern, R. (eds), *GATS 2000: New Directions in Services Trade Liberalization*, Center for Business and Government, Harvard University and the Brookings Institution Press, Washington DC.

Discussant — Alison Burrows

Department of Foreign Affairs and Trade

The Department of Foreign Affairs and Trade strongly endorses the project on impediments to services trade by the Australian National University and the Productivity Commission. I wish to confine my comments to making some observations on related developments in the working groups under the WTO Council for Trade in Services, particularly:

- the WTO Committee on Specific Commitments (CSC), which deals with scheduling of market access and national treatment; and
- the Working Party on Domestic Regulation (WPDR).

These comments cover classification issues, currently being considered by the CSC, and how Members are proposing to address the ‘sins of omission’. Here, the Basic Telecommunications Reference Paper (NGBT 1996) is a useful model. They also cover how Australia and some other Members are proposing to address the ‘sins of commission’ in the WPDR by developing a necessity test.

D14.1 Categories of impediments

Tony Warren’s paper addresses how market access and national treatment are defined in the GATS. As the paper points out, Article XVI (market access) of the GATS does not specifically define market access, but it does include an exhaustive list of measures considered to be limitations on market access. On the other hand, Article XVII (national treatment) defines national treatment as ‘no less favourable treatment ...’, but there is no illustrative list.

The CSC has recently been considering precisely this issue. Although Members are currently showing a reluctance to revise the guidelines on market access, they are more disposed to revisit national treatment. The CSC is considering a draft list of national treatment restrictions. The examples are closely inspired by existing entries in the schedules, and the idea is to retain them as an illustrative, rather than an exhaustive list, particularly since a specific definition of national treatment already exists.

The draft list covers:

- subsidy measures;
- tax measures;
- other financial measures (eg fees and charges);
- nationality requirements;

-
- residency requirements;
 - licensing and qualification requirements;
 - registration requirements;
 - authorisation requirements;
 - technology transfer and training requirements;
 - local content requirements; and
 - ownership of property and land.

The CSC is working to clarify the obscure nature of the schedules and to avoid misscheduling. On the basis of this work, the way Tony Warren's paper defines the terms 'market access' and 'national treatment' is inconsistent with what the CSC is developing. Tony Warren defines market access as non-discrimination between incumbents, and national treatment as non-discrimination between domestic and foreign service suppliers. Although the concept of discrimination is embedded in national treatment, market access cannot be defined entirely in terms of non-discrimination. There are measures that limit all suppliers from getting access to the market. Snape (1998) noted:

... they can embrace not only restrictive trade practices but legislative barriers to entry to an activity, as well as government regulation of access to essential facilities where 'natural monopoly' elements are present.

D14.2 Addressing the sins of omission

Turning now to the central thrust of the paper, the sins of omission, it is important to recall that the GATS is based on the concept that Members have a right to regulate. The preamble (the 'chapeau') of the GATS says:

Recognising the right of Members to regulate, and to introduce new regulations, on the supply of services within their territories in order to meet national policy objectives.

The problem of failing to regulate is an issue attracting increasing attention in the preparatory stage of the service negotiations. The Australian Government has focused its efforts in this area in seeking broader support for the Basic Telecommunications Reference Paper (NGBT 1996), described by Aaditya Mattoo as one of the central achievements of the GATS. For example, Australia requests all acceding countries to the WTO to endorse the principles of the Reference Paper.

The Australian Government is also exploring whether the concepts contained in the Reference Paper could be applied to other sectors. Those concepts are:

- competitive safeguards;

-
- interconnection;
 - universal services;
 - public availability of licensing criteria;
 - independent regulators; and
 - allocation and use of scarce resources.

These concepts could have application in other sectors, especially those sectors having networks or segments operating under natural monopoly conditions. The OECD Secretariat has recently completed an excellent paper on interconnection and on the possible application of the Reference Paper to other services — energy, transport, postal and courier, and environmental (OECD 1999).

D14.3 Addressing the sins of commission

Australia has also been active in addressing ‘sins of commission’ in the WTO Working Party on Domestic Regulation (WPDR). Last year, Australia introduced into the WPDR the idea of a necessity test for regulation, based on the necessity test used in the Agreement on the Application of Sanitary and Phytosanitary (SPS) Measures. Article VI.4 (Domestic Regulation) of the GATS calls for the establishment of a necessity test at a later date, and it is worth noting that the GATS is something of a work-in-progress. At the moment, such a necessity test would only apply to measures relating to qualification requirements and procedures, technical standards, and licensing requirements, as Members are resisting broadening the coverage originally envisaged.

Australia is proposing text for the necessity test, which would refer to taking account of international standards and the need for an independent regulator. The central point of Australia’s proposed text would be:

... a measure is not more trade restrictive than required unless there is another measure reasonably available, taking into account technical and economic feasibility, that achieves a legitimate policy objective and is significantly less restrictive to trade.

What this means is that binding disciplines would be developed to ensure that members do in fact use the least trade restrictive measure. This could be tested in dispute settlement. Australia is not proposing, however, that other Members should be able to question whether or not specific policy objectives are ‘legitimate’. Australia does not want to undermine the principle of leaving it up to governments to determine their policy objectives.

D14.4 Conclusion

Finally, in a paper presented at the World Services Congress in Atlanta last year, Feketekuty (1999) argued that:

The negotiation of international agreements on regulatory issues is thus an exercise in balancing the benefits and costs of economies of scale and scope in regulation. Liberalising trade inevitably creates the need for rules that limit national flexibility. It is all the more necessary to remind oneself regularly of the need to minimise the intrusion into national decision, just as much as it is desirable to limit regulatory intrusion into individual choice by consumers and enterprises.

Although it is clear that there will be no rapid conclusion to the services negotiations, Members' consideration of regulatory issues in the WTO is evolving relatively rapidly. Because of the failure to launch a round at Seattle, Members are turning to technical issues, including domestic regulation. Although we are in a preparatory phase, there are signs that at the end of the first phase of negotiations, by March 2001, Members will have developed new frameworks for considering both the sins of omission and the sins of commission.

References

- Feketekuty, G. 1999, Assessing the WTO General Agreement on Trade in Services and improving the GATS architecture, Paper presented at the World Services Congress, Atlanta, Georgia, 1–3 November.
- NGBT (Negotiating Group on Basic Telecommunications) 1996, 'Reference paper on telecommunication regulation', WTO, Geneva.
- OECD (Organization for Economic Cooperation and Development) 1999, 'Assessing barriers to trade in services: potential interconnection disciplines and their relevance to specific sectors', TD/TC/WP(99)44/REV1, OECD Secretariat, Paris.
- Snape, R. 1998, 'Reaching effective agreements covering services', in Krueger, A. (ed.), *The WTO as an International Organization*, University of Chicago Press, Chicago and London.

Discussant — *Will Martin*

World Bank

The distinction between sins of omission and commission in Tony Warren's paper provides an interesting and useful framework. It also relates very well to the previous discussion about the distinction between regulations oriented towards internalising externalities, reducing market power or overcoming information deficiencies.

A useful way of thinking about the welfare impacts of government interventions is with a balance of trade function. This is just the consumer expenditure function [$e(p, X, v, u)$], minus the revenue function [$r(p, X, v)$] and the price distortion [$(p - p^*)'$] times the volume of trade [$z_p(p, X, v, u)$].

$$B = e(p, X, v, u) - r(p, X, v) - (p - p^*)' z_p(p, X, v, u)$$

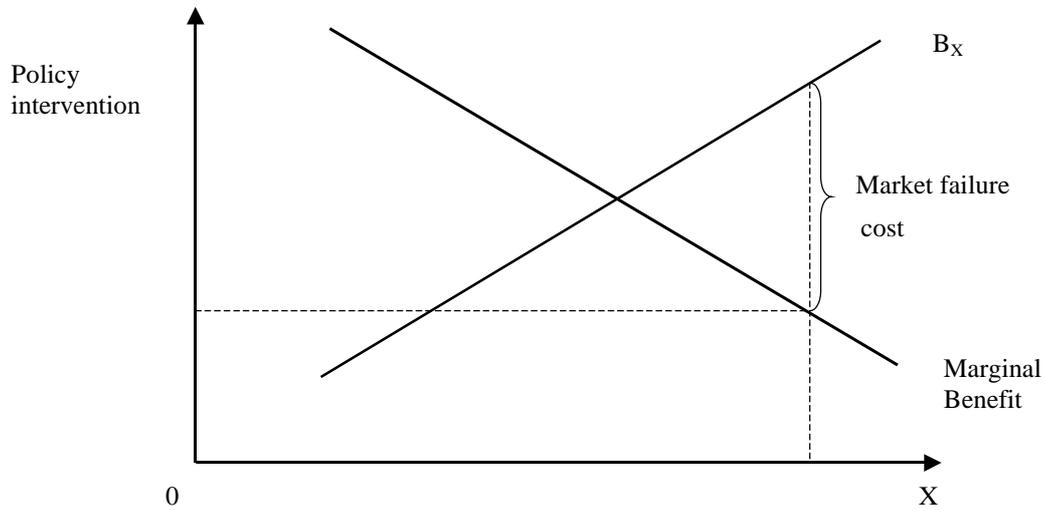
where $z_p = e_p - r_p$.

All of the elements of welfare change are encapsulated in this framework. The welfare impact of a price distortion, as the domestic price (p) moves away from the world price (p^*), is a standard convex function of the size of the distortion.

This framework can be easily extended to incorporate externalities or information deficiencies. A vector X can be used to represent factors, such as the stock of carbon dioxide, that affect the consumer expenditure required for the representative consumer to achieve a given level of utility or the income generated from production. The marginal cost of the pollution is the derivative of the balance of trade function with respect to the variable X . In the case of information deficiencies, this derivative might show the marginal cost of generating the information required by a particular regulatory regime.

This framework allows comparison between the marginal cost and the marginal value or the demand for the externality or public information generated by the regulatory policy. If there is a demand for pollutants that is created by the desire to drive motor cars, weighing up the marginal costs and the marginal value associated with a given level of air pollutants shows the gap between the demand price and the supply price. This gap gives a measure of the cost of market failure (figure 14.1). The marginal benefit of improved information can likewise be associated with a downward-sloping schedule of marginal benefits that can, in turn, be associated with the marginal cost schedule for providing this information.

Figure 14.1 Comparing marginal cost and marginal benefit of an intervention



B_x is the marginal cost of pollution, or of information generated by a policy intervention.

A government policy that seeks to address a distortion has a cost associated with it — a sin of either omission or commission — unless it exactly matches marginal benefits and costs. Identifying the gap between demand and supply of the benefit in question gives a framework to formalise and weigh up the balance between marginal costs and benefits of regulations. This cost-benefit framework is fundamentally what underlies a serious evaluation of regulatory interventions.

General discussion

The discussion focused on:

- the definition of market access and national treatment;
- necessity and proportionality tests;
- GATS coverage of subsidies; and
- the need for generic rules covering both goods and services.

Definition of market access and national treatment

In response to the discussant's comment on the definition of market access, Tony Warren responded that the only way to operationalise the concept of market access was to define it as discrimination against entrants or discrimination towards incumbents. Discrimination had to be the key. Otherwise it would be difficult to operationalise the concept without having an exhaustive list specifying what regulations were in and what were not.

One participant agreed that the market access provisions within the GATS had been a cause for concern, because they went beyond border measures towards domestic regulation. The national treatment provisions had a clear place in the GATS, because they addressed discrimination between domestic and foreign suppliers. Market access provisions existed because it was argued that there was no use having national treatment if foreign suppliers could not get into the market, and, in this sense, barriers to entry were barriers to trade. Market access provisions to provide for generic access to essential facilities were appropriate, but care was needed in going beyond this point.

Necessity and proportionality tests

One reason the 'necessity test' was quite significant in the multilateral context was that it did not question a country's objectives or the level to which the objectives should be achieved, but only disciplined the choice of instrument. But in the national context, it might be possible to go even further and apply a 'proportionality test', where the strength of the instrument had to be in some sense proportional to the problem being solved. In some ways, governments already applied such a test. For example, in principle, the quality of doctors could be improved by forcing them to train for ten years rather than five, but an assessment was typically made as to the contribution that the extra training would make to improving the state of health services. In a domestic context, there was the possibility of balancing the marginal

benefit against the marginal cost in a way that the multilateral system today could not.

Alison Burrows pointed out that some countries had introduced the idea of a proportionality test in a multilateral context, but it was hard to see whether the proposal would be accepted because it had only recently been introduced.

Tony Warren said that governments should be interested in measuring necessity and proportionality in order to ensure appropriate domestic regulation. He was wary of the WTO having a role because it was difficult enough to get appropriate domestic law, without requiring international agreement. He also thought that domestic courts would do a better job than the very political dispute resolution system in Geneva in resolving domestic disputes.

Aaditya Mattoo noted that the necessity test should be seen as an add-on discipline, supplementing the non-discriminatory rules that already existed in national treatment provisions. Not even the existing rules — which prohibit *de facto* discrimination — could be made operational without applying some sort of necessity test.

The crucial role of the WTO today was to prevent countries discriminating between like services and like suppliers, but the way likeness was determined had everything to do with domestic regulation. If a foreign doctor was like a domestic doctor, a country could not impose an even slight additional burden on the foreign doctor. If the foreign doctor was not like a domestic doctor, as determined by the domestic regulator, then national treatment simply did not apply.

Taking away the right of regulators to impose even a slightly additional burden on foreign providers was not politically sustainable. But giving regulators the full freedom to determine who would be treated like a domestic supplier would be equivalent to abandoning all multilateral rules. Between these two extremes, the necessity test provided a happy medium. Regulators should determine the extent to which a foreign provider differed from a domestic provider in terms of training, but should then subject the qualification or licensing requirement to a simple test — was the regulation necessary to achieve the objective of the policy?

The proposed necessity test nicely balanced permissiveness with regard to objectives against discipline regarding instruments. It might not always give an optimal outcome, but would at least screen out the more egregious qualification and training requirements, for example, requiring doctors to go back to the beginning of medical school.

GATS coverage of subsidies

One participant expressed concern about the GATS covering subsidies (Article XV). In a services context, a possible question was whether subsidies to education were a subsidy to the subsequent delivery of services by the trained professional. Nothing had been done about subsidies in the GATS, and perhaps that was just as well, because the issues were complex and there were other issues that were more important.

The need for generic rules covering both goods and services

The same participant stressed the importance of focusing on generic policies and rules covering many areas, including goods, wherever possible. While a multilateral investment agreement was not possible in the foreseeable future, the existing investment rules hardly covered any goods at all. Investment by service firms was covered under GATS, but investment by manufacturing firms was virtually not covered. Such inconsistent regulation across goods and services was inefficient.

There was also inconsistency between the GATT and the GATS with respect to recognition of standards and mutual recognition. There were many such areas where it was possible to develop generic rules and apply the principles of good regulation across goods and services, rather than making changes or developing new rules that may cause as much trouble as they solved.

15 Modelling the liberalisation of services

Philippa Dee, Kevin Hanslow and Tien Phamduc[†]

Productivity Commission

15.1 Introduction

Examining the effects of eliminating barriers to trade in services, while important in itself, can give some insights into the likely effects of achieving better regulation of services. This is because barriers to services trade can be thought of as a subset of inappropriate domestic regulation. So while the paper focuses on trade, it also notes the prospects for extending the methodology to look at domestic regulation more generally.

Barriers to trade in services are unlike tariffs, the main traditional tool of trade policy. They are typically regulatory barriers, rather than explicit taxes. They need not discriminate against foreigners. Indeed, barriers to market access are often designed to protect incumbent firms from *any* new entry, be it by domestic or foreign firms. And barriers to services trade are not restricted to affecting the *output* of services firms. One particularly important category of barriers to services trade — restrictions on foreign direct investment by service firms — affects the use of primary factors. These restrictions are recognised in the General Agreement on Trade in Services (GATS) under the WTO, since this agreement recognises commercial presence as one of the modes by which services are traded.

To date, few papers of either a theoretical or empirical nature have reviewed all these aspects of barriers to services trade. A few theoretical papers in the late 1980s examined some of the important characteristics of services, including knowledge intensity (eg Markusen 1989, Melvin 1989). This characteristic also featured in subsequent analysis of goods trade under imperfect competition (eg Grossman and

[†] The views expressed in this paper are those of the authors and do not necessarily reflect those of the Productivity Commission. The authors are grateful for comments from Kym Anderson, Chang-Tai Hsieh, Fukunari Kimura, Anne Krueger and Will Martin.

Helpman 1991). But those early theoretical papers did not look at the nature of barriers to services trade. Recently, a few empirical papers have examined the effects of removing barriers to trade in services. Many of these have failed to take account of barriers to commercial presence as an important category of barrier to trade in services (Brown et al. 1995, Brown, Deardorff and Stern 1996, Hertel 1999, Nagarajan 1999). One seminal paper by Petri (1997) introduced a treatment of barriers to foreign direct investment in the services sector, but failed to take into account barriers on the other modes of service delivery. And all empirical papers have suffered from a dearth of convincing empirical estimates of the incidence of barriers to services trade.

A recent empirical paper by Dee and Hanslow (2000) sought to analyse the effects of removing barriers to services trade in a more comprehensive fashion. The barriers included non-discriminatory barriers to market access, as well as discriminatory restrictions on national treatment. They included barriers to commercial presence as well as barriers to the other modes of service delivery. The focus of that paper was to compare the gains from liberalising services trade with the gains from removing all post-Uruguay barriers to trade in agriculture and manufacturing. The paper also compared the gains from the total removal of barriers to services trade with the gains from several alternative approaches to partial liberalisation. It identified significant second-best problems with some approaches to partial liberalisation.

The purpose of this paper is to look more deeply at that analysis of services trade liberalisation to assess the extent to which the adjustment mechanisms from traditional trade theory are still relevant in a more realistic model of services trade liberalisation. In the process, the analysis examines whether and how the benefits of services trade liberalisation are passed on to other sectors in the economy. Thus, the analysis tries to open up the ‘black box’ of what is a rather complex general equilibrium model of services trade, in order to gain insights into the sectoral results from that model in terms of more simple textbook treatments of trade policy analysis.

The structure of the paper is as follows. It first describes the model used — a multi-sector, multi-regional computable general equilibrium model of world trade and investment. The theoretical structure of the model covers both FDI and portfolio investment. The model’s database contains estimates of FDI stocks and the activities of FDI firms, each on a bilateral basis. Thus the model recognises that both goods and services can be delivered via FDI as well as by conventional trade. The paper then looks at the size of the barriers to trade in services, and the cost impost they impose on other sectors of the economy. This analysis uses the first of a comprehensive new set of estimates of barriers to services trade. To understand the

general equilibrium effects of removing these barriers, the effects on each sector in selected economies are built up from a more restricted, partial equilibrium multi-country model. To this partial model are gradually added the resource constraints and income linkages associated with general equilibrium. The paper then briefly summarises the implications of services trade liberalisation for regional incomes. Finally, the paper identifies areas for further research.

15.2 The FTAP model

The model is a version of GTAP (Hertel 1997) with foreign direct investment, known as FTAP. The treatment of FDI follows closely the pioneering work of Petri (1997). FTAP also incorporates increasing returns to scale and large-group monopolistic competition in all sectors. This follows Francois, McDonald and Nordstrom (1995), among others, who adopted this treatment for manufacturing and resource sectors, and Brown et al. (1995) and Markusen, Rutherford and Tarr (1999), who used similar treatments for services. Finally, FTAP makes provision for capital accumulation and international borrowing and lending. This uses a treatment of international (portfolio) capital mobility developed by McDougall (1993), and recently incorporated into GTAP by Verikios and Hanslow (1999). FTAP is implemented using the GEMPACK software suite (Harrison and Pearson 1996). Its structure is documented fully in Hanslow, Phamduc and Verikios (1999). The model and its documentation are available at the Productivity Commission web site at <http://www.pc.gov.au>.

Theoretical structure

FTAP takes the standard GTAP framework as a description of the *location* of economic activity, and then disaggregates this by *ownership*. For example, each industry located in Australia comprises Australian owned firms, along with US, Japanese and other multinationals. Each of these firm ownership *types* is modelled as making its own independent choice of inputs to production, according to standard GTAP theory. And each firm type has its own sales structure.

On the purchasing side, agents in each economy make choices among the products or services of each firm type, distinguished by both ownership and location, and then among the individual (and symmetric) firms of a given type. Thus, the model recognises the firm-level product differentiation associated with monopolistic competition. Firms choose among intermediate inputs and investment goods, while households and governments choose among final goods and services.

Agents are assumed to choose first among products or services from domestic or foreign locations, with a CES elasticity of substitution of 5. They then choose among particular foreign locations, and among ownership categories in a particular location, both with a CES elasticity of substitution of 10. Finally, they choose among the individual firms of a particular ownership and location, with a CES elasticity of substitution of 15. With firm-level product differentiation, agents benefit from having more firms to choose among, because it is more likely that they can find a product or service suited to their particular needs. Capitalising on this, Francois, McDonald and Nordstrom (1995) show that the choice among individual firms can be modelled in a conventional model of firm types (not firms) by allowing a productivity improvement whenever the output of a particular firm type (and hence the number of individual firms in it) expands. But because the substitutability among individual firms is assumed here to be very high, the incremental gain from greater variety is not very great and this productivity enhancing effect is not particularly strong (the elasticity of productivity with respect to output¹ is $1/15 = 0.0667$).

The first two choices, among domestic and foreign locations, are identical to the choices in the original GTAP model. They have been parameterised using values, 5 and 10, that are roughly twice the standard GTAP Armington elasticities. Two reasons can be given for doubling the standard elasticities. One is that only with such elasticities can GTAP successfully reproduce historical changes in trade patterns (Gehlhar 1997). The other is that higher elasticities accord better with notions of firm level product differentiation. Further calibration of the model to historical data using methods of maximum entropy (eg Liu, Arndt and Hertel 2000) may provide a feasible means of refining the above estimates of firm-level substitution possibilities in the future.

The order of the first three choices, among locations and then among ownership categories, is the opposite of the order adopted by Petri (1997). The current treatment assumes that from an Australian perspective, for example, a US multinational located in Australia is a closer substitute for a Australian owned firm than it is for a US firm located in the United States. Petri's treatment assumes that US owned firms are closer substitutes for each other than for Australian firms, irrespective of location.

There are two reasons for preferring the current treatment.

¹ The equivalent elasticity of productivity with respect to *inputs* is $0.0667/(1 - 0.0667) = 0.0714$, where this latter concept is used by Francois, McDonald and Nordstrom (1995). The elasticities of productivity with respect to output and inputs are not equal because of the assumption of increasing returns to scale. Another reason that scale effects are not strong is that, with this nested structure, the economies of scale are regional rather than global.

The first is that Petri's treatment produces a model in which multilateral liberalisation of tariffs on manufactured goods produces large economic welfare losses, for most individual economies and for the world as a whole — an uncomfortable result at odds with conventional trade theory. The reason for the result is spelt out in more detail in Dee and Hanslow (2000).

The second reason for preferring the current treatment is that, in many instances, it accords better with reality. One of the distinguishing characteristics of services is that they are tailored each time to meet the needs of the individual consumer. Another characteristic is that they are often delivered face to face, sometimes making commercial presence (through FDI) the only viable means of trade. These taken together mean that service firms in a given location, irrespective of ownership, will tailor their services to meet local tastes and requirements, and thus appear to be close substitutes, as in the current treatment.

While the demand for the output of firms distinguished by ownership and location is determined as above, the supply of FDI is determined by the same imperfect transformation among types of wealth as in Petri (1997). Investors in each economy first divide their wealth between 'bonds' (which can be thought of as any instrument of portfolio investment), real physical capital, and land and natural resources in their country of residence. This choice is governed by a CET semi-elasticity of 1, meaning that a one percentage point increase in the rate of return on real physical capital, for example, would increase the ratio of real physical capital to bond holdings by one per cent. A bond is a bond, irrespective of who issues it, implying perfect international arbitrage of rates of return on bonds. However, capital in different locations is seen as different things. Investors next choose the industry sector in which they invest (with a CET semi-elasticity of 1.2). They next choose whether to invest at home or overseas in their chosen sector (with a CET semi-elasticity of 1.3). Finally, they choose a particular overseas region in which to invest (with a CET semi-elasticity of 1.4).

The less than perfect transformation among different forms of wealth can be justified as reflecting some combination of risk aversion and less than perfect information. It is important to note, however, that while the measure of economic welfare in FTAP currently recognises the positive income contribution that FDI can make, it does not discount that for any costs associated with risk taking, given risk aversion. This is an important qualification to the current results, and will be the subject of further research.

While the chosen CET parameters at each 'node' of the nesting structure may appear low, the number of nests means that choices at the final level (across destinations of FDI) are actually very flexible. For example, it can be shown that, holding total wealth fixed but allowing all other adjustments across asset types and

locations to take place, the implied semi-elasticity of transformation between foreign destinations can easily reach 20, and be as high as 60. The variation across regions in these implied elasticities comes about because of the different initial shares of assets in various regional portfolios.

The choice of CET parameters at each ‘node’ was determined partly by this consideration of what they implied for the final elasticities, holding only total wealth constant. They were also chosen so that this version of FTAP gave results that were broadly comparable to an earlier version of GTAP with imperfect international (portfolio) capital mobility, for experiments involving the complete liberalisation of agricultural and manufacturing protection (Verikios and Hanslow 1999). Imperfect capital mobility was also a feature of the GTAP-based examination of APEC liberalisation by Dee, Geisler and Watts (1996) and Dee, Hardin and Schuele (1998). These parameters thus provide a familiar starting point from which refinements could be made in the future, possibly based on methods of maximum entropy.

In one respect, however, the current version of FTAP does differ from previous versions of GTAP with imperfect capital mobility. The GTAP variants assumed that capital was perfectly mobile across sectors, whereas FTAP has less than perfect sectoral mobility. Furthermore, the choice of sector is relatively early in the nesting structure, so that the implied elasticities guiding choice of sector, holding only total wealth constant, are relatively low (eg 1.2 in the United States). As a result, FTAP tends to exhibit behaviour where resources move less readily between sectors in a given region, but more readily across regions in a given sector, although the differences are not dramatic. The current treatment is consistent with the idea that the knowledge capital often required to succeed in foreign direct investment, despite the difficulties of language and distance, is likely to be sector-specific.

Petri’s model assumed that total wealth in each region was fixed. In FTAP, while regional endowments of land and natural resources are fixed (and held solely by each region’s residents), regional capital stocks can accumulate over time, and net bond holdings of each region can adjust to help finance the accumulation of domestic and foreign capital by each region’s investors. The treatment of capital accumulation follows the original treatment of McDougall (1993), and was also used by Verikios and Hanslow (1999), Dee, Geisler and Watts (1996) and Dee, Hardin and Schuele (1998).

With this treatment of capital accumulation, FTAP provides a long-run snapshot view of the impact of trade liberalisation, ten years after it has occurred. To the extent that liberalisation leads to changes in regional incomes and savings, this will be reflected in changes to the capital stocks that investors in each region will have been able to accumulate. As noted, investors in each region are not restricted to

their own savings pool in order to finance capital investment. They may also issue bonds to help with that investment, but only according to their own preferences about capital versus bond holding, and only according to the willingness of others to accept the additional bonds.

Model database

The starting point for FTAP's database was not the standard GTAP database, since this includes measures of trade and investment barriers that are still to be eliminated under the Uruguay Round agreement. Instead, the starting point was an updated version of the GTAP database, following a simulation in which the barriers yet to be eliminated under the Uruguay Round had been removed. Such a database was provided by the work of Verikios and Hanslow (1999), under their assumption of less than perfect capital mobility.

FDI data

The Petri treatment of FDI requires the addition of data on bilateral FDI stocks, and on the activity levels and cost and sales structures of FDI firms. The methods used to estimate such data were similar to those of Petri. APEC (1995) and UNCTAD (1994) provided limited data on FDI stocks by source, destination and sector. These data were fleshed out to provide a full bilateral matrix of FDI stocks by source, destination and sector, using RAS methods (Welsh and Strzelecki 2000). The resulting estimates are summarised in Dee and Hanslow (2000). The data were collected (and the model implemented) for 19 regions and three broad sectors. The three sectors — primary (agriculture, resources and processed food), secondary (other manufacturing), and tertiary (services) — correspond broadly to the three areas of potential trade negotiation in a new trade round. The intention is to use similar methods to produce a model with greater sectoral detail in the future.

The FDI stock data were used in turn to generate estimates of the output levels of FDI firms. To do this, capital income flows were estimated by multiplying the FDI stocks by rates of return. These capital rentals were then grossed up to get an output estimate for FDI firms, using capital rental to output ratios from the GTAP database. Again, the resulting estimates are summarised in Dee and Hanslow (2000). A possible future refinement would be to use additional information on the ratio of value added to output from US and Japanese data on the activities of offshore affiliates (eg Baldwin and Kimura 1998, Kimura and Baldwin 1998).

The detailed cost and sales structures of FDI firms were assumed to be the same as for locally owned firms, and were obtained by pro-rating the GTAP database. A

subject for future research would be to make use of information on the true cost and sales structures of FDI firms, again using available US and Japanese data on the activities of offshore affiliates.

Estimates of barriers to services trade

Estimates of existing barriers to services trade were injected into the model's database, using the techniques of Malcolm (1998). The process is documented in Hanslow et al. (2000).

The estimates of barriers to services trade were the first of a comprehensive new set of estimates documented in Findlay and Warren (2000). As noted in chapter 14 of this conference volume, the general methodology of these studies is as follows.

- Qualitative information on barriers to services trade is converted to a quantitative index measure of trade restrictiveness, based on coverage and some initial judgements about the relative restrictiveness of the different sorts of restrictions.
- An econometric model is developed to measure the determinants of the economic performance (eg price, profit margin, cost, or quantity) of service firms in a given sector in different countries, taking account of all the factors that economic theory would suggest relevant, including the index measure of trade restrictiveness.
- The economic model is used to estimate the determinants of economic performance. Wherever possible, the components of the trade restrictiveness index are entered separately so that the econometrics can reveal something about the relative weights attached to the separate components.²
- The results of the econometrics are used to calculate the effect of trade restrictions on performance. Where necessary, quantity or profit effects are converted to price or cost effects.

When considering how this methodology can be generalised to assess domestic regulation more generally, it is important to note that there is no presumption that complete deregulation would be the best regulatory stance. In some cases, regulation deemed to be appropriate to meet a well-defined objective could be simply omitted from the 'restrictiveness' index. In other cases, a regulatory best-practice benchmark could be identified, and an index used to score deviations from that benchmark. Indeed, one of the challenges to characterising regulatory regimes and tracing their effects on economic outcomes is to identify such regulatory

² This is not possible where there is high multicollinearity between the various components, or where there is a lack of in-sample variation in some of the components.

benchmarks. Many of the papers in this conference volume make significant steps in that direction.

In assessing barriers to services trade, estimates of barriers to trade in banking services were taken from Kalirajan et al. (2000), and estimates of barriers to trade in telecommunications services were taken from Warren (2000). The rates can be taken as indicative of post-Uruguay rates, since the Uruguay Round established the architecture for services trade negotiations, but did not achieve much in the way of services trade liberalisation (Hoekman 1995).

For modelling purposes, the barrier estimates were decomposed according to a two-by-two classification.

- The GATS framework distinguishes four modes of service delivery — via commercial presence, cross-border supply, consumption abroad, and the presence of natural persons. Accordingly, the FTAP model distinguishes barriers to establishment from barriers to ongoing operation. This is similar to the distinction between commercial presence and other modes of delivery, since barriers to establishment are a subset of barriers to commercial presence. Barriers to establishment are modelled as taxes on the movement of capital. Barriers to ongoing operation are modelled as taxes on the output of the service-providing firms.
- The GATS framework also distinguishes restrictions on market access from restrictions on national treatment. As noted above, the former are restrictions on entry, be it by locally owned or foreign owned firms. In the FTAP model, they are treated as non-discriminatory. Restrictions on national treatment mean that foreign owned firms are treated less favourably than domestic firms are. These restrictions are treated as discriminatory.

The decomposition of trade barriers into this two-by-two classification follows the classifications used by Kalirajan et al. (2000) and Warren (2000). Table 15.1 shows how they classified barriers to trade in banking and telecommunications services. Note that in the banking sector, prudential regulations were not counted as trade barriers, nor included in the restrictiveness index. This was based on the recognition that they were designed to address a genuine market failure, and the judgement that they were generally implemented in an appropriate fashion to that end. It is also consistent with the so-called ‘prudential carve-out’ allowed for in the GATS.

Table 15.1 **Classifying barriers to trade in banking and telecommunications services**

	<i>Non-discriminatory barriers to market access</i>	<i>Discriminatory derogations from national treatment</i>
Barriers to establishment		
<i>Banking</i>	Are there restrictions on the number of bank licences?	Are there restrictions on the number of foreign bank licences? Are there restrictions on foreign equity investment or requirements for foreigners to enter through a joint venture with a domestic bank? Are there restrictions on the permanent movement of people?
<i>Telecommunications</i>	One measure of restriction is actual number of competitors in fixed and mobile markets. Is there an enforced monopoly, partial competition or full competition in various fixed line markets and mobile market? What percentage of the incumbent fixed or mobile operator is privatised?	What percentage of foreign investment is allowed in competitive carriers?
Barriers to ongoing operation		
<i>Banking</i>	Are there general restrictions on raising funds, lending, providing other lines of business, or expanding the number of banking outlets?	Are foreign banks restricted in raising funds, lending, providing other lines of business, or expanding the number of banking outlets? Are there restrictions on the proportion of foreigners on the board of directors? Are there restrictions on the temporary movement of people?
<i>Telecommunications</i>	Are there restrictions on leased lines or private networks? Are there restrictions on third party resale? Are there restrictions on connection of leased lines and private networks to the public switched telephone network?	Are there restrictions on callback services?

Source: McGuire and Schuele (2000) and Warren (2000).

Note also that in the banking study, horizontal (ie not sector-specific) restrictions on the permanent movement of people were counted as a barrier to establishment, and hence modelled as a barrier to the movement of capital. More properly, these restrictions should be modelled as a barrier to the movement of labour, but so far FTAP does not allow for international labour mobility. Similarly, horizontal restrictions on the temporary movement of people were counted as a barrier to ongoing operation, affecting both offshore affiliates and services delivered via ‘cross-border’ trade, where the latter is broadly defined to include services delivered via the temporary movement of the consumer or the producer. In reality, the barriers affecting true cross-border trade are sufficiently different from those affecting trade involving temporary movement to warrant modelling them separately. These are areas for further research.

A simple average of the estimated price effects of barriers to trade in banking and telecommunications was taken as being typical of most services — all of the GTAP service categories of trade and transport and finance, business and recreational services, and half of public administration and defence, education and health. The remainder of public administration and defence, education and health, along with electricity, water and gas, construction, and ownership of dwellings were assumed to be strictly non-traded (note that engineering services are part of business services, not construction). The resulting average estimates of barriers to trade in the tertiary sector would have been about 50 to 100 per cent bigger, had the banking and telecommunications estimates been taken as indicative of the whole of the service sector. A topic of future research is to use the next version of the GTAP database, which will have more service sector detail, to model barriers to each service separately, thus overcoming the extreme arbitrariness of these assumptions on coverage.

The resulting structure of post-Uruguay barriers to trade in services is summarised in table 15.2. Barriers to trade in primary (agricultural, resource and processed food) and secondary (manufacturing) products are also shown for comparison purposes. Barriers to primary products are represented via a combination of taxes on imports, and subsidies (shown in table 15.2 as negative taxes) on exports and output. Unfortunately, at FTAP’s three sector level of aggregation, the actual taxes on primary exports and output are a combination of subsidies used for protective purposes, and taxes (eg excises on alcohol and tobacco) used for revenue raising. (While the average taxes on primary output are not shown in table 15.2, they are all relatively small, and mostly positive.) In future, using a database with greater sectoral detail will reduce the problems associated with ‘aggregation bias’.

Table 15.2 Tax equivalents of post-Uruguay barriers to trade and investment
(per cent)

	<i>Imports</i>		<i>Exports</i>		<i>Domestic output</i>	<i>Foreign affiliates' output</i>	<i>Domestic capital</i>	<i>Foreign affiliates' capital</i>
	<i>Pri</i>	<i>Sec</i>	<i>Pri</i>	<i>Ter</i>	<i>Ter</i>	<i>Ter</i>	<i>Ter</i>	<i>Ter</i>
Australia	1.69	7.30	0.65	4.81	0.00	0.69	0.62	14.79
NZ	1.16	4.51	-3.25	3.78	0.00	0.67	0.41	4.18
Japan	16.19	1.81	-8.12	4.41	3.59	4.75	0.33	3.01
Korea	12.95	6.61	-1.22	4.57	5.11	6.78	1.91	22.01
Indonesia	4.40	6.71	0.00	4.68	13.23	28.11	22.69	68.06
Malaysia	21.18	5.97	6.68	4.50	3.58	10.20	15.35	37.58
Philippines	16.16	18.51	-0.10	4.80	8.38	22.65	7.40	54.28
Singapore	3.22	0.56	0.01	4.70	3.40	8.32	2.42	24.50
Thailand	12.12	14.81	-16.98	4.14	4.69	13.36	12.16	36.49
China	8.92	28.45	5.13	4.08	18.75	36.40	123.46	250.66
Hong Kong	0.00	0.00	0.00	9.91	1.39	2.36	1.35	5.41
Taiwan	27.31	5.63	-1.82	4.35	2.88	4.90	1.90	19.19
Canada	3.57	1.40	-0.43	3.54	0.25	1.67	0.53	6.11
USA	1.29	2.24	-0.02	4.26	0.07	1.08	0.00	3.83
Mexico	-1.50	2.99	1.89	5.23	2.17	5.59	0.68	12.99
Chile	6.76	10.26	0.02	4.36	2.97	4.11	14.15	20.36
R. Cairns	3.82	13.39	6.30	4.49	0.98	5.55	7.19	19.45
EU	3.17	1.13	-2.33	4.72	0.10	1.31	1.33	6.49
R. World	15.94	13.67	0.59	4.95	4.89	13.92	39.07	86.97

Pri: primary sector (agriculture, resources and processed food); Sec: secondary sector (manufacturing); Ter: tertiary sector (services). NZ: New Zealand; USA: United States of America; R. Cairns: Rest of Cairns group — Brazil, Argentina, Colombia and Uruguay; EU: European Union; R. World: Rest of the World.

Source: FTAP model database.

In the service sector, as noted above, barriers to establishment have been modelled as taxes on capital. Barriers to ongoing operation may affect either FDI firms or those supplying via the other modes, and have been modelled as taxes on the output of locally-based firms (either domestic or foreign owned), and taxes of the same size on the exports of firms supplying via the other modes, respectively. The estimates of export taxes on services in the fourth column of table 15.2 are trade weighted averages of the taxes on exports to particular destinations, where these are equal in turn to the taxes on foreign affiliates' output in the destination region, shown in the sixth column. These are modelled as taxes in the exporting region, rather than as tariffs in the importing region, to allow the rents created by the barriers to be retained in the exporting region. The issue of rents is addressed in more detail shortly.

The model also distinguishes restrictions on market access from restrictions on national treatment. The taxes on domestic capital and domestic output in table 15.2 represent the effects of restrictions on market access (affecting establishment and ongoing operation, respectively). The taxes on the capital and output of foreign affiliates are higher than the corresponding taxes on domestic firms, because they represent the effects of restrictions on both market access and national treatment.

The estimates in table 15.2 indicate that barriers to trade in services are generally at least as large as those are on agricultural and manufactured products. Most economies have at least some significant barriers to trade in services. The only regions where barriers are low across the board are New Zealand, Japan, Hong Kong, Canada, the United States and the European Union. But this statement should be heavily qualified, because it is based only on estimates of barriers to banking and telecommunications.

Barriers to trade in services have been modelled as tax equivalents that generate rents — a mark-up of price over cost — rather than as things that raise costs above what they might otherwise have been (Hertel 1999). This decision was based on the way in which the price impacts of barriers to trade in banking and telecommunications services were measured. Kalirajan et al. (2000) measured the effects of trade restrictions on the net interest margins of banks, a direct measure of banks' mark-up of price over cost.³ Warren (2000) measured the effects of trade restrictions on the quantities of telecommunications services delivered, and these were converted to price impacts using an estimate of the elasticity of demand for telecommunications services. Thus, Warren's estimates did not provide direct evidence of a mark-up of price over cost, but the relative profitability of telecommunications companies in many countries suggests that some element of rent may exist. By contrast, there is evidence that trade restrictions in sectors such as aviation raise costs (Johnson et al. 2000). As estimates of the effects of trade barriers in these sectors are incorporated into the model, it will be appropriate to treat some restrictions as cost-raising rather than as rent-creating.

One important implication of the current treatment is that welfare gains from liberalising trade in services are likely to be understated, perhaps significantly. If trade restrictions create rents, then the allocative efficiency gains from trade liberalisation are the 'triangle' gains associated with using a given quantum of resources more efficiently. By contrast, if trade restrictions raise costs, the gains from trade liberalisation include 'rectangle' gains (qualified by general equilibrium

³ Net interest margins — a measure of the difference between borrowing and lending rates of interest — can also be thought of as the 'price' of financial intermediation services. The econometric model used to test the significance of barriers to trade in banking services was developed from an economic model of financial intermediation.

effects) from lower costs, equivalent to a larger effective quantum of resources for productive use.

Because barriers to services trade appear to be significant, and because they have been modelled as taxes, the rents they generate will be significant. A key issue is whether those rents should be modelled as being retained by incumbent firms, appropriated by governments via taxation, or passed from one country to another by transfer pricing or other mechanisms. In FTAP, the rents on outputs have been modelled as accruing to the selling region, and those on capital have been modelled as accruing to the region of ownership, once the government in the region of location has taxed them at its general property income tax rate. Despite this, the asset choices of investors are modelled as being driven by pre-tax rates of return. This is because many economies, in the developed world at least, have primarily destination-based tax systems. For example, if tax credits are granted for taxes paid overseas, investors are ultimately taxed on *all* income at the owning region's tax rate. Although such tax credits have not been modelled explicitly, their effect has been captured by having investors respond to relative pre-tax rates of return. Nevertheless, investor choices are also assumed to be determined by rates of return excluding any abnormal rent component. Investors would like to supply an amount of capital consistent with rates of return including abnormal rents, but are prevented from doing so by barriers to investment. The amount of capital actually supplied is, therefore, that amount that investors would like to supply at rates of return excluding abnormal rents.

Thus a portion of the rent associated with barriers to services trade is assumed to remain in the region of location in the form of property income tax revenue, while the remainder accrues to the region of ownership. Thus liberalisation of services trade could have significant income effects in both home and host regions as these rents are gradually eliminated. Dee and Hanslow (2000) show in detail how significant these effects are, relative to the allocative efficiency effects and other effects normally associated with trade liberalisation.

A final point to note is that the model's database does not contain estimates of barriers to investment in agriculture and manufacturing, even though they are likely to be significant. It is unlikely that a new trade round would include negotiations on them. Nevertheless, their omission will affect the model's estimates of the effects of liberalisation elsewhere, and the results need to be qualified accordingly.

15.3 The cost impact of barriers to trade in services

Table 15.2 shows that the direct ‘tax equivalents’ of barriers to trade in services are often significant, compared with the trade barriers expected to remain in agriculture and manufacturing after full implementation of the Uruguay Round. It also shows that barriers to services trade tend to be much higher in developing than in developed economies.

A priori, this does not mean that the service sectors in developing economies would suffer most from services trade liberalisation. Barriers to services trade are unlike tariffs, so there are two key mechanisms by which the service sectors in developing countries could expand following services trade liberalisation.

- Not all services trade barriers discriminate against foreign service suppliers, so the service sector could expand because of new domestic entry.
- Some services trade barriers restrict inward FDI, so the service sector could expand because of new foreign entry.

These mechanisms could be sufficient to offset the traditional mechanisms by which a protected sector can be harmed by removal of protection.

- Some services barriers discriminate against foreign services delivered cross-border, so the service sector could contract in the face of additional import competition.
- Services trade liberalisation may benefit downstream using industries, and the service sector may lose out in the competition for domestic resources (eg labour).

Figure 15.1 examines the extent to which downstream using industries are likely to benefit from services trade liberalisation. It shows the direct and indirect cost impost of domestic barriers to trade in services on all sectors in selected typical model regions, as calculated from the FTAP model database.

In general terms, the figure shows the direct and indirect input requirements needed to produce a unit of final demand in each sector. For example, a unit of processed food (a primary activity) sold to households might require inputs of unprocessed food (another primary activity), as well as packaging materials from the secondary sector. The packaging materials might again require inputs from forestry (a primary activity), along with electricity from the tertiary sector. Each of these direct and indirect inputs would have their own requirements for labour, capital, fixed factors (land and natural resources) and imported inputs, and these can be added up. Where the cost of the direct and indirect inputs is inflated by taxes, the direct and indirect tax contributions can also be calculated.

Figure 15.1 Direct and indirect input requirements per unit of final demand

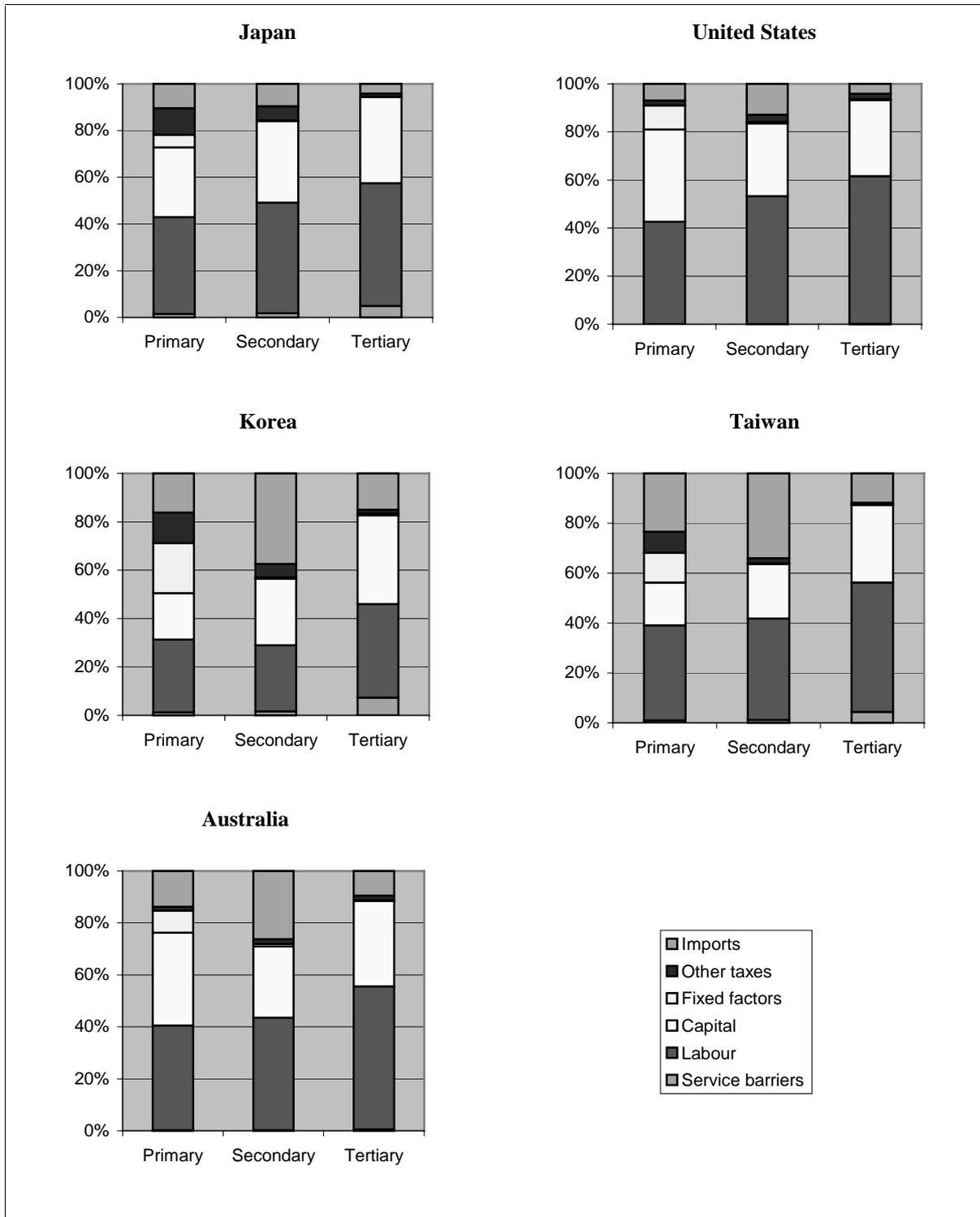
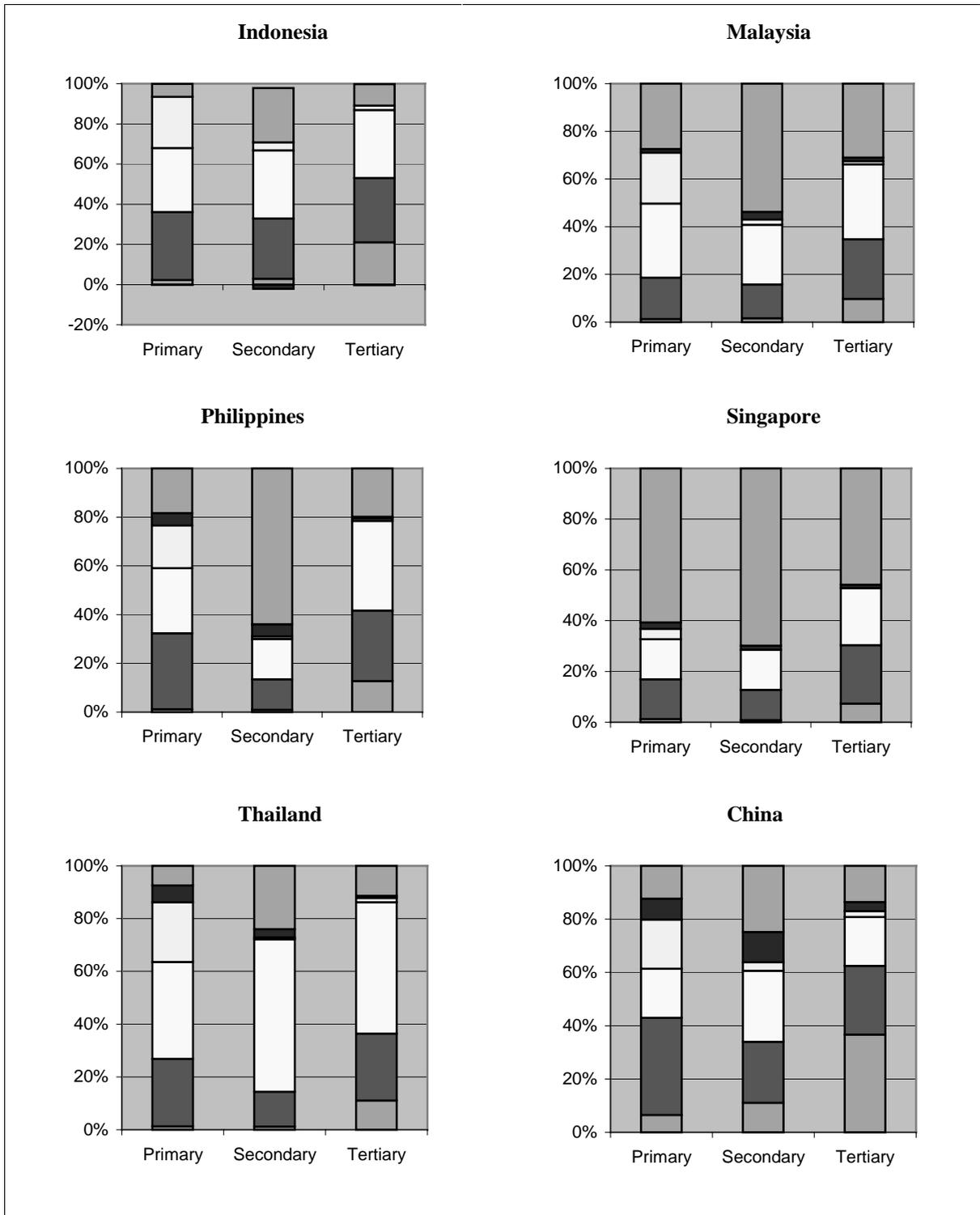


Figure 15.1 (continued)



Thus, the direct and indirect cost impost of domestic barriers to services trade has been calculated by adding together the following:

- the output and capital taxes on direct and indirect services inputs, where those taxes represent the effects of domestic barriers to commercial presence (both establishment and ongoing operation); and
- the export taxes (in the source region) falling on direct and indirect imported inputs, where these export taxes represent the effects of domestic barriers to cross-border services trade (where the term ‘cross-border’ is interpreted loosely to include services traded via the temporary movement of the producer or consumer).

All other domestic taxes are collected in the contribution of ‘Other taxes’, and all other taxes on imports (primarily tariffs) are included with the contribution of ‘Imports’.

Figure 15.1 shows that, in every region shown, the greatest unit cost impost from services trade barriers falls on the service sector itself. This reflects two factors. First, the service sector experiences a direct taxing effect, whereas in other sectors the burden is indirect, through the higher cost of service inputs. Second, this effect is reinforced by the fact that in both developed and developing economies, the service sector itself tends to have a higher direct services input requirement than any other sector. Although other sectors may need service inputs, the greatest intensity of use of services is within the service sector itself. Thus, as will be seen, the benefits of services trade liberalisation in many economies are concentrated within the service sector. This result is contrary to the normal effects of tariff removal, where the benefits are typically concentrated in other sectors.

Another feature of figure 15.1 is that in the economies with the highest per capita incomes (Japan, United States, Korea, Taiwan and Australia), the cost impost of domestic services trade barriers on other sectors is minimal. Although these economies tend to be more service-dependent, in terms of having higher direct service input requirements, their domestic barriers to services trade are also relatively low.

Somewhat surprisingly, in the economies with the lowest per capita incomes (Indonesia, Malaysia, Philippines, Singapore, Thailand and China), the cost impost of domestic services trade barriers on other sectors is not much greater. Only in China, where services trade barriers are particularly high, does the cost impost on other sectors approach 10 per cent.⁴

⁴ The cost impost is estimated to be particularly high in China because its telecommunications market is particularly restrictive. When estimates of services barriers are incorporated for a

By showing the cost impost of only domestic barriers to trade in services, figure 15.1 understates the potential first-round impact of multilateral liberalisation of services trade. When barriers are removed globally, not only will domestic goods and services be cheaper, but so too will goods and services available from other economies. This benefit is likely to be significant in the highly import-intensive economies such as Korea, Taiwan, Malaysia, the Philippines and Singapore. And since the trade and transport services used to ship goods internationally will also be cheaper, there will be an additional cost reduction effect not captured in figure 15.1.

15.4 The sectoral effects of removing barriers to services trade

Partial equilibrium effects

A useful way to understand the sectoral effects of removing barriers to services trade is to start with a partial equilibrium framework, and to gradually add the economywide constraints that distinguish a general from a partial equilibrium approach. This is a very useful technique of analysis developed by Hertel (1997).

An initial partial equilibrium model is obtained by ‘turning off’ the following parts of FTAP.

- Factor supply constraints. Each sector in each region can get all the labour and capital it needs at the going wage or rental price. Thus, the secondary and tertiary sectors in each region have horizontal supply curves (which nevertheless move downward as services barriers are removed). The primary sector continues to have an upward sloping supply curve because fixed factors (land and natural resources) are still treated as being in fixed supply in each economy.
- Income linkages. Irrespective of what is projected to happen to factor prices and other variables, the model’s measure of welfare is held fixed in each region. This ‘equivalent variation’ (EV) is essentially a measure of net national product, or the real income accruing to the residents of each economy. In general equilibrium, it is affected not just by the amount of activity generated within a region, but also by net foreign interest and dividend payments associated with foreign borrowing and lending and with FDI.
- The endogenous productivity changes associated with a love of variety.

broader range of services than banking and telecommunications, the overall cost imposts could differ from those shown here. Not only could the overall impost in China be lower, but the impost in developed countries could also be higher (since banking and telecommunications happen to be sectors in which developed countries are particularly liberal).

In partial equilibrium, all the demand side substitution possibilities of the full FTAP model are still in operation. Thus, for example, the demand for the output of the secondary sector in a region will depend on:

- how the cost (and hence price) of its output changes relative to the cost (and price) of output of secondary sectors in other economies, and how consumers and users in each region substitute between domestic and various imported sources of secondary output as a result of those relative price changes;
- how the cost (and hence price) of its output changes relative to the average price (across sources) of primary and tertiary output, and how domestic consumers and the government substitute between the outputs of these different sectors as a result of these relative price changes;⁵ and
- what the secondary input requirements are per unit of output in other sectors, and whether those other sectors are expanding or contracting.

Thus, even in the partial equilibrium model, the richness of substitution possibilities and inter-industry linkages on the demand side make for a rather complicated story.

Because real incomes in each economy are assumed to be fixed, it would be expected that, unless substitution effects dominate, the demand for, and hence output of, a commodity or service should increase whenever services trade liberalisation reduces its price. And the only sector in which services trade liberalisation would conceivably *not* reduce the price is the primary sector, where the return to the fixed factor could conceivably be bid up. Thus, the presumption is that services trade liberalisation should reduce prices and increase output. Where this does not occur, it must generally be as a result of substitution effects.

Within the service sector itself, prices fall and output rises in the ASEAN economies and China (table 15.3). Note that although the prices of domestic services fall in these economies, the prices of imported services fall by significantly more. Thus, substitution towards imports in these economies might suggest that domestic services output should fall. But offsetting this is an increase in exports of services from these economies. In the services sector, the price of a service import in the destination country can fall by significantly more than its output price in the exporting country. This is primarily because services trade liberalisation involves removing the ‘export tax’ equivalent of barriers to cross-border trade imposed by the destination country. So although domestic services in the ASEAN region and China are disadvantaged relative to imports at home, when those same services are exported, their prices compare favourably with service exports from most other

⁵ In FTAP, as in GTAP, consumers and government are the only agents to substitute directly among different commodities. For intermediate and investment usage, different commodities (aggregated across sources) are used in fixed proportions.

regions. (This is indicated indirectly by the fact that the domestic output price of services falls by more in ASEAN and China than in the other regions.) Thus, the services output expansion in ASEAN and China is primarily an export story.

In the higher per capita income economies (Korea, Taiwan and Australia), services output falls, despite a reduction the domestic price, because of substitution towards imports. This is in accordance with the relative price movements shown in table 15.3.

The declines in the output of the secondary sector in Japan, the United States and Australia are because of substitution towards imports, especially in intermediate usage. For the other higher income economies shown (Korea and Taiwan), the prices of domestic secondary output do not change greatly relative to secondary import prices, so the secondary output expansions in these economies are primarily an export story. In ASEAN and China, the secondary output expansions are because of both increased exports and substitution away from imports.

Table 15.3 Partial equilibrium effects on selected regions of removing global barriers to trade in services, by sector
(per cent)

	<i>Primary</i>			<i>Secondary</i>			<i>Tertiary</i>		
	<i>Q</i>	<i>Pd</i>	<i>Pm</i>	<i>Q</i>	<i>Pd</i>	<i>Pm</i>	<i>Q</i>	<i>Pd</i>	<i>Pm</i>
Japan	-0.3	-2.4	-0.9	-1.4	-2.6	-3.9	-3.4	-2.1	-21.9
United States	-1.3	-0.7	-1.3	-7.3	-0.6	-3.3	-4.3	-0.4	-13.6
Korea	-0.1	-1.9	-1.0	2.3	-2.9	-3.1	-2.3	-3.5	-16.3
Taiwan	1.2	-1.2	-0.9	2.9	-2.3	-2.9	-4.5	-2.5	-14.7
Australia	1.9	-0.5	-1.2	-7.2	-1.3	-3.0	-5.5	-1.2	-12.1
Indonesia	2.7	-0.5	-0.9	8.8	-3.6	-3.1	13.4	-12.1	-30.6
Malaysia	4.2	0.5	-0.9	6.9	-3.3	-2.9	0.3	-8.2	-21.3
Philippines	1.2	-1.1	-0.9	2.9	-2.9	-2.9	8.1	-7.5	-27.9
Singapore	18.2	-1.6	-0.9	8.9	-3.5	-3.0	1.9	-6.7	-19.6
Thailand	3.1	0.6	-1.3	-5.3	-1.9	-3.0	0.3	-7.6	-21.9
China	36.6	18.1	-1.2	132.0	-10.2	-2.6	245.2	-27.9	-31.9

Q: domestic output quantity, Pd: domestic price; and Pm: import price.

Source: FTAP model projections with partial equilibrium closure.

While in the secondary and tertiary sectors the results are driven primarily by substitution among different sources of each commodity, in the primary sector it is possible to see the effects of households in each region substituting among different commodities. This explains the slight falls in the output of the primary sector in Japan and Korea. In these economies, the prices of imported services fall significantly more than the prices of any other final commodity. Households tend to substitute towards imported services and away from everything else. Thus primary output in these economies falls, despite the price of domestic primary output falling by more than its import price.

In the United States, the effect on primary sector output of households switching away from the primary sector in general is reinforced by substitution (in relative terms) towards primary imports.

In Taiwan, Australia, Singapore and Thailand, the expansion of the primary sector is primarily an export story. (The landed cif price of Thai primary exports falls, despite a slight increase in the domestic output price, because of cheaper international trade and transport services.) This can be confirmed by looking at more detailed model results not shown in table 15.3.

In Indonesia, Malaysia, the Philippines and China, the switch by households away from the primary sector in general is offset by increased intermediate input demand, and some increase in export demand, for primary sector output. The increased intermediate demand occurs despite an adverse relative price movement against imports (in all but the Philippines), because of inter-industry linkages between the primary sector and the downstream secondary and tertiary sectors.

In summary, multilateral liberalisation of services trade reduces domestic costs and prices across all economies, and the partial equilibrium sectoral effects are of three types.

- In economies such as Japan and the United States, where initial domestic services barriers are particularly low, domestic prices do not fall by much, and substitution towards cheaper imports leads to a reduction in output in all sectors of the economy. Real income can remain constant, however, because of the cheaper imports.
- At the other extreme, in the economies of the ASEAN region and China, where initial domestic services barriers are relatively high, domestic prices tend to fall significantly, and output in (almost) all sectors of these economies expands.
- In between are the economies of Korea, Taiwan and Australia, where initial domestic services barriers are moderate, but where all sectors are more trade exposed than in Japan and the United States. Thus, while the service sectors in

these economies may not benefit from services trade liberalisation, at least some of their other sectors benefit from cheaper domestic and imported inputs and thus gain an advantage on export markets.

General equilibrium effects

The partial equilibrium results of table 15.3 assumed that each sector in each economy could get any additional labour and capital at the going wage or rental price. The results also ignored the income implications of services trade liberalisation.

In table 15.4, these effects are gradually reintroduced into the model. The first column reproduces the partial equilibrium results from table 15.3. In the second column, factor supply constraints are imposed. As in textbook models, aggregate supplies of capital and labour are assumed to be fixed, and these factors are treated as being perfectly mobile within each sector of the economy. In the third column, sectoral capital stocks are assumed to take the values they would in the full general equilibrium model. Thus, not only do aggregate capital stocks change in each economy, but capital is no longer perfectly mobile across sectors. Finally, the full general equilibrium results are presented. These incorporate not only the primary factor behaviour of the full general equilibrium model, but also the associated income effects (including the net foreign income flows associated with foreign direct investment).

In broad terms, the imposition of factor supply constraints is the single most important step in taking the partial equilibrium sectoral results towards their general equilibrium values.

Even with factor supply constraints, the results for the tertiary sector in each region are qualitatively quite close to the partial equilibrium results:

- the service sectors in Japan and the United States are still smaller than in the absence of services trade liberalisation;
- the service sectors in most other high income economies are also projected to decline (Korea is the exception); and
- the service sectors in the ASEAN region and China still gain from services trade liberalisation.

But now the wage/rental ratios in each economy adjust to ensure that the induced output changes in other sectors do not lead to a violation of the overall factor supply constraints. Thus, the output of the primary and secondary sectors in Japan, the United States, Taiwan and Australia is now projected to rise to counteract the

decline in their service sectors. And in the ASEAN region, China and Korea, output in many of the primary and secondary sectors is now projected to decline to offset the expansion of their service sectors.

The results in the third column hint at the complexity of the capital supply story in the full FTAP model. Even though services trade liberalisation involves removing taxes on service sector capital, it is not always the case that cautious investors would invest more in those service sectors than they would if they viewed investment in any sector as being equally desirable (consistent with perfect sectoral capital mobility). In Japan, the United States, Korea, Taiwan and Australia, service sector capital stocks are larger than in the ‘fixed factors’ case, but in the other economies they are smaller. This demonstrates how the capital supply behaviour in the FTAP model plays an important role in relocating capital across regions within a sector, as opposed to the textbook treatment of capital allocation across sectors within a region.

The final column of table 15.4 incorporates the FTAP model’s income linkages — real income in each region is no longer constant, but reflects the induced changes in factor prices and international capital movements. Dee and Hanslow (2000) demonstrate that such income effects are crucial to the welfare implications of liberalising trade in services, as will be seen shortly. But table 15.4 shows that these income effects do not have strong additional effects on the sectoral distribution of gains from services trade liberalisation.

General equilibrium models are often regarded as ‘black boxes’, with there being little chance of understanding what is inside. The above analysis suggests that because the structure of barriers to services trade is complex, the hardest part about understanding the effects of multilateral liberalisation of services trade is understanding what happens in partial equilibrium.

The partial equilibrium results help to demonstrate how liberalisation of services trade can differ from tariff removal. Barriers to services trade affect domestic new entrants as well as foreign suppliers, and the sector to benefit most in output terms from liberalisation can often be the service sector itself. This result is robust in the transition from partial to general equilibrium.

Table 15.4 Partial and general equilibrium effects on sectoral output in selected regions of removing global barriers to trade in services
(per cent)

		<i>Full PE</i>	<i>Fixed factors</i>	<i>Capital as in GE</i>	<i>Full GE</i>
Japan	Primary	-0.3	0.2	-0.3	-0.4
	Secondary	-1.4	0.9	-0.5	-0.3
	Tertiary	-3.4	-0.3	0.2	0.1
United States	Primary	-1.3	2.4	0.4	0.6
	Secondary	-7.3	1.9	0.2	0.6
	Tertiary	-4.3	-0.7	-0.2	-0.4
Korea ^a	Primary	-0.1	-0.5	-0.7	-0.8
	Secondary	2.3	-0.4	-1.4	-1.6
	Tertiary	-2.3	0.3	1.0	1.1
Taiwan ^a	Primary	1.2	0.4	-0.1	0.1
	Secondary	2.9	2.5	0.2	1.0
	Tertiary	-4.5	-1.1	0.1	-0.2
Australia ^a	Primary	1.9	5.1	1.0	1.2
	Secondary	-7.2	3.0	1.2	1.0
	Tertiary	-5.5	-1.1	-0.3	-0.3
Indonesia ^a	Primary	2.7	-3.4	0.3	0.3
	Secondary	8.8	-9.7	2.5	2.6
	Tertiary	13.4	8.9	8.5	9.2
Malaysia ^a	Primary	4.2	-1.1	0.0	0.1
	Secondary	6.9	-1.8	0.2	0.1
	Tertiary	0.3	3.3	1.5	1.5
Philippines ^a	Primary	1.2	-3.0	-1.9	-1.9
	Secondary	2.9	-6.6	-2.6	-3.6
	Tertiary	8.1	5.6	2.3	2.5
Singapore	Primary	18.2	3.2	-4.0	-4.0
	Secondary	8.9	-4.2	-5.6	-6.6
	Tertiary	1.9	4.5	0.6	1.0
Thailand ^a	Primary	3.1	0.7	-0.2	-0.1
	Secondary	-5.3	-8.0	-0.7	-0.8
	Tertiary	0.3	8.1	1.3	1.3
China ^a	Primary	36.6	-6.5	-1.1	-0.2
	Secondary	132.0	-16.7	4.5	2.5
	Tertiary	245.2	43.7	28.7	32.5

PE: partial equilibrium; GE: general equilibrium.

^a Aggregate capital stock is projected to increase in the general equilibrium closure.

Source: FTAP model projections with partial and general equilibrium closures.

General equilibrium welfare effects

The first column of table 15.5 summarises the effects of full liberalisation of services trade on economic well-being in selected model regions (Dee and Hanslow 2000 present results for all model regions). As in GTAP, the measure of economic well-being is the equivalent variation — essentially a measure of the change in real income in each region, where the deflator is an index of the prices of household consumption, government consumption and national saving. But for FTAP, the relevant measure of national income is net national product — the income accruing to the residents of a region — rather than net domestic product — the income generated within the borders of a region. Thus, net domestic product is adjusted for the income earned on outward FDI, net of the income repatriated overseas from inward FDI, plus the income from net bond holdings.

Table 15.5 **Welfare effects of full multilateral liberalisation of services trade**
Absolute change in \$US million

	<i>Equivalent variation</i>	<i>Contribution to EV from change in^a</i>			
		<i>Endowments</i>	<i>Real FDI stocks</i>	<i>Real bond holding</i>	<i>Rents on FDI capital and output</i>
Japan	4 130	-1 030	3 120	-2 978	-8 730
United States	-1 809	-5 713	2 665	1 708	-6 716
Korea	1 886	438	-5	39	123
Taiwan	-142	312	378	-583	-423
Australia	2 098	58	0	4	495
Indonesia	2 470	7 158	-541	-4 519	530
Malaysia	1 015	367	-103	-168	585
Philippines	1 236	164	-91	47	214
Singapore	-247	-1 071	-198	-108	1 049
Thailand	1 698	305	-24	-393	486
China	90 869	52 164	-12 649	-5 776	12 849

^a This table reports a subset of the factors contributing to the change in the equivalent variation. For example, the table omits the contribution made by changes in terms of trade. As a result, the four right-hand columns do not add up to the equivalent variation reported.

Source: FTAP model projections with general equilibrium closure.

Three of the selected economies are projected to have incomes lower than otherwise as a result of full multilateral liberalisation of services trade — the United States, Taiwan and Singapore. Dee and Hanslow (2000) show that in each case, the losses from multilateral liberalisation of services trade would be more than offset by income gains accruing from multilateral liberalisation of trade in agriculture and

manufacturing. Nevertheless, the source of the income losses from multilateral liberalisation of services trade warrants further investigation, especially for the United States, where the losses are projected to be significant.

Dee and Hanslow show that for agricultural and manufacturing liberalisation, the welfare results are dominated by two things — the contribution of improvements in allocative efficiency, and the contribution of induced changes in the terms of trade. The model's regions are projected to experience positive income gains, or in a few cases small losses, as a result of these effects.

For services liberalisation, changes in foreign direct investment patterns contribute several additional effects. First, FDI can lead to an expansion or contraction in the capital stock located within a region, leading to a positive or negative contribution to income generated within a region from this change in national endowments. Second, it can lead to changes in net FDI and net lending positions, with consequent changes in net foreign income flows accruing to residents. Third, it can induce changes in the returns earned on those net foreign asset holdings. An important example here is changes in the rents earned on foreign direct investment.

The second column of table 15.5 shows the contribution to real income from changes in real capital endowments. Generally, if capital endowments are higher than otherwise, real GDP will be higher than otherwise, and vice versa.⁶ A major reason why Singapore is projected to lose slightly from services trade liberalisation is because its capital stock is projected to be lower than otherwise.

But a lower capital stock located domestically need not reduce incomes for domestic residents. Earnings from higher outward FDI and higher lending abroad could offset it. The third column of table 15.5 shows the contribution to residents' real income from changes in real FDI stocks. The fourth column shows the contribution from changes in real bond holdings. Both columns also help to indicate the way in which changes in capital endowments are financed.

For example, Japan's capital stock is lower than otherwise, but it has a big increase in outward FDI. In fact, it also borrows (a negative change in bond holding) in order to finance its outward FDI. By contrast, China's increase in capital endowments comes partly from a large increase in inward FDI, and partly from additional foreign borrowing. Thus the large projected increase in China's service sector output and exports, noted above, comes as much from an expansion in foreign-owned service firms located in China as it does from an expansion in Chinese-owned service firms. The United States is projected to have a smaller capital

⁶ For a few regions, real GDP can be higher than otherwise, even if endowments are lower than otherwise, because the endowments are used more efficiently.

endowment than otherwise, but this is offset to some extent by an increase in outward FDI and increased lending to other regions.

For a few regions, real incomes are affected not so much by changes in net asset positions, but by changes in returns on those assets. Although the details are not shown in table 15.5, Taiwan is projected to lose slightly from services trade liberalisation, primarily because in the FTAP database it is a net creditor economy and is adversely affected by a small induced fall in real interest rates.⁷

A further source of change in asset returns is the change in rents generated by barriers to services trade. The last column of table 15.5 shows the income contribution to recipient countries of changes in the rents accruing to foreign direct investment, as barriers to services trade are eliminated. What is striking is the loss of rents to the main providers of outward FDI — Japan and the United States. In fact, the loss of rents to US incumbent multinationals is more than sufficient to explain its overall projected income loss from multilateral liberalisation of services trade. But note that this result is sensitive to the assumption that all barriers to services trade are rent-creating rather than cost-raising.

Generally, while induced changes in capital stocks — both those located domestically and those owned abroad — do not appear to play a major role in explaining the effects of multilateral services trade liberalisation on sectoral output, they play a major role in explaining the effects on real regional incomes. Barriers to services trade affect capital movements as well as the output of services firms, so services trade liberalisation can have a significant effect on the regional location and ownership of capital. The flow-on effects to regional incomes demonstrate another way in which liberalisation of services trade can differ from tariff removal.

15.5 Agenda for further research

Much of the research agenda for further development of the FTAP model has been outlined already. It involves continuing to obtain estimates of the price impacts of barriers to services trade along the lines outlined in Findlay and Warren (2000), both for additional sectors and for additional modes of service delivery within a sector. The methodologies should in the process reveal whether the barriers are rent-creating or cost-raising. Such methods could also be used to estimate the price impact of barriers to foreign direct investment in agriculture and manufacturing. More sectoral detail needs to be incorporated into FTAP, to model the barriers to each service separately. More research is required to obtain more realistic output

⁷ Interest rates fall primarily because of an assumption that government savings *rates* are held constant. Growing revenues and savings *levels* therefore allow some government debt retirement.

estimates and cost and sales structures for FDI firms and, if possible, a realistic initial allocation of rents. And the welfare measure in FTAP needs to be amended to take account of the costs of risk taking, given risk aversion.

In addition, some of the simplifying assumptions made during the original development of FTAP should now be relaxed, and the sensitivity of the results to these assumptions tested. One such assumption was the uniformity of behavioural parameters across sectors and regions. While this reflected a deliberate research strategy, its importance could be tested using systematic sensitivity analysis (Arndt and Pearson 1996). The importance of data issues (eg the initial distribution of rents) and theoretical issues (eg investor behaviour) could also be explored.

But there is also scope for much more work using simple analytical models of services trade that better incorporate the features of services and the nature of the barriers to their trade. Insights of the sort available in Markusen, Rutherford and Hunter (1995), for example, provide invaluable guidance to those attempting to build better empirical models of FDI and services trade.

References

- APEC (Asia Pacific Economic Cooperation) 1995, *Foreign Direct Investment and APEC Economic Integration*, APEC Economic Committee, Singapore.
- Arndt, C. and Pearson, K. 1996, 'How to carry out systematic sensitivity analysis via Gossan Quadrature and GEMPACK', GTAP Technical Paper no. 3, Centre for Global Trade Analysis, Purdue University.
- Baldwin, R.E. and Kimura, F. 1998, 'Measuring US international goods and services transactions', in Baldwin, R.E., Lipsey, R. and Richardson, J.D. (eds), *Geography and Ownership as Bases for Economic Accounting*, University of Chicago Press, Chicago, pp. 9–48.
- Brown, D.K., Deardorff, A.V., Fox, A.K. and Stern, R.M. 1995, 'Computational analysis of goods and services liberalisation in the Uruguay Round', in Martin, W. and Winters, L.A. (eds), *The Uruguay Round and the Developing Economies*, Discussion Paper no. 307, World Bank, Washington DC, pp. 365–80.
- , ——— and Stern, R. 1996, 'Modelling multilateral trade liberalisation in services', *Asia Pacific Economic Review*, vol. 2, no. 1, pp. 21–34.
- Dee, P., Geisler, C. and Watts, G. 1996, *The Impact of APEC's Free Trade Commitment*, Industry Commission Staff Information Paper, AGPS, Canberra.

-
- and Hanslow, K. 2000, *Multilateral Liberalisation of Services Trade*, Productivity Commission Staff Research Paper, AusInfo, Canberra.
- , Hardin, A. and Schuele, M. 1998, *APEC Early Voluntary Sectoral Liberalisation*, Productivity Commission Staff Research Paper, AusInfo, Canberra.
- Findlay, C. and Warren, T. (eds) 2000, *Impediments to Trade in Services: Measurement and Policy Implications*, Routledge, London and New York (forthcoming).
- Francois, J.F., McDonald, B. and Nordstrom, H. 1995, 'Assessing the Uruguay Round', in Martin, W. and Winters, L.A. (eds), *The Uruguay Round and the Developing Economies*, Discussion Paper no. 307, World Bank, Washington DC, pp. 117–214.
- Gehlhar, M. 1997, 'Historical analysis of growth and trade patterns in the Pacific Rim: an evaluation of the GTAP framework', in Hertel, T. (ed.), *Global Trade Analysis: Modelling and Applications*, Cambridge University Press, Cambridge, pp. 349–63.
- Grossman, G. and Helpman, E. 1991, *Innovation and Growth in the Global Economy*, MIT Press, Cambridge, Massachusetts.
- Hanslow, K., Phamduc, T. and Verikios, G. 1999, 'The structure of the FTAP model', Research Memorandum Cat. no. MC 58, Productivity Commission, Canberra, December.
- , —, — and Welsh, A. 2000, 'Incorporating barriers to services trade into the FTAP database', Research Memorandum Cat. no. MC 59, Productivity Commission, Canberra, April.
- Harrison, J.W. and Pearson, K.R. 1996, 'Computing solutions for large general equilibrium models using GEMPACK', *Computational Economics*, vol. 9, no. 2, pp. 83–127.
- Hertel, T. 1997, *Global Trade Analysis: Modelling and Applications*, Cambridge University Press, Cambridge.
- 1999, Potential gains from reducing trade barriers in manufacturing, services and agriculture, Paper presented at the 24th Annual Economic Policy Conference, Federal Reserve Bank of St. Louis, 21–22 October.
- Hoekman, B. 1995, 'Assessing the General Agreement on Trade in Services', in Martin, W. and Winters, L.A. (eds), *The Uruguay Round and the Developing Economies*, Discussion Paper no. 307, World Bank, Washington DC, pp. 327–64.

-
- Johnson, M., Gregan, T., Belin, P. and Gentle, G. 2000, 'Modelling the impact of regulatory reform', Chapter 7 in Findlay, C. and Warren, T. (eds), *Impediments to Trade in Services: Measurement and Policy Implications*, Routledge, London and New York (forthcoming).
- Kalirajan, K., McGuire, G., Nguyen-Hong, D. and Schuele, M. 2000, 'The price impact of restrictions on banking services', Chapter 13 in Findlay, C. and Warren, T. (eds), *Impediments to Trade in Services: Measurement and Policy Implications*, Routledge, London and New York (forthcoming).
- Kimura, F. and Baldwin, R.E. 1998, 'Application of a nationality-adjusted net sales and value-added framework: the case of Japan', in Baldwin, R.E., Lipsey, R. and Richardson, J.D. (eds), *Geography and Ownership as Bases for Economic Accounting*, University of Chicago Press, Chicago, pp. 49–82.
- Liu, J., Arndt, C. and Hertel, T. 2000, Estimating trade elasticities for GTAP: a maximum entropy approach, Paper presented at the Third Annual Conference in Global Economic Analysis, Melbourne, 26–29 June.
- McDougall, R. 1993, 'Incorporating international capital mobility into SALTER', SALTER Working Paper no. 21, Industry Commission, Canberra.
- McGuire, G. and Schuele, M. 2000, 'Restrictiveness of international trade in banking services', Chapter 12 in Findlay, C. and Warren, T. (eds), *Impediments to Trade in Services: Measurement and Policy Implications*, Routledge, London and New York (forthcoming).
- Malcolm, G. 1998, 'Adjusting tax rates in the GTAP data base', GTAP Technical Paper no. 12, Center for Global Trade Analysis, Purdue University.
- Markusen, J. 1989, 'Trade in producer services and in other specialised intermediate inputs', *American Economic Review*, vol. 79, no. 1, pp. 85–95.
- , Rutherford, T.F. and Hunter, L. 1995, 'Trade liberalisation in a multinational dominated industry', *Journal of International Economics*, vol. 38, no. 1–2, pp. 95–117.
- , ——— and Tarr, D. 1999, Foreign direct investment in services and the domestic market for expertise, Paper presented at Second Annual Conference on Global Economic Analysis, Denmark, 20–22 June.
- Melvin, J. 1989, 'Trade in producer services: a Heckscher-Ohlin approach', *Journal of Political Economy*, vol. 97, no. 5, pp. 1180–96.
- Nagarajan, N. 1999, 'The millennium round: an economic appraisal', Economic Papers no. 139, Directorate-General for Economic and Financial Affairs, European Commission, November.

Petri, P.A. 1997, Foreign direct investment in a computable general equilibrium framework, Paper prepared for the conference 'Making APEC work: Economic Challenges and Policy Alternatives', Keio University, Tokyo, 13–14 March.

UNCTAD (United Nations Conference on Trade and Development) 1994, *World Investment Directory, Foreign Direct Investment, Legal Framework and Corporate Data*, vol. IV, *Latin America and the Caribbean*, Division on Transnational Corporations and Investment, United Nations, New York.

Verikios, G. and Hanslow, K. 1999, Modelling the effects of implementing the Uruguay Round: a comparison using the GTAP model under alternative treatments of international capital mobility, Paper presented at Second Annual Conference on Global Economic Analysis, Denmark, 20–22 June.

Warren, T. 2000, 'The impact on output of impediments to trade and investment in telecommunications services', Chapter 6 in Findlay, C. and Warren, T. (eds), *Impediments to Trade in Services: Measurement and Policy Implications*, Routledge, London and New York (forthcoming).

Welsh, A. and Strzelecki, A. 2000, Estimating domestic and foreign returns to capital for the FTAP model, Research Memorandum Cat. no. MC 60, Productivity Commission, Melbourne, May.

Discussant — *Kym Anderson*

University of Adelaide

This is a crucially important paper that pushes beyond the frontier of global modelling of goods trade to show how policies affecting international trade and investment in services might also be modelled in the same comparative static framework. Despite the greater complexity that this economywide perspective involves, the paper is a pleasure to read because it walks the reader step-by-step through less-familiar territory towards the empirical results, while providing many thought-provoking ideas along the way. The paper will be of interest, not only to trade economists seeking to model the effects of GATS commitments, but also to analysts focussed on domestic services regulation needing an open-economy view, given that those markets are increasingly being influenced by international forces as deregulation and globalisation proceed.

The paper has three distinct sections that outline the model the authors have developed, the way they incorporate barriers to trade and investment in services, and the results they generate on the sectoral effects of removing those barriers.

D15.1 The model

The first section of the paper details the modifications that the authors have made to the standard computable general equilibrium model of the global economy known as GTAP (after Purdue University's Global Trade Analysis Project). The GTAP model is a successor to the *SALTER* model developed by Productivity Commission staff in the late 1980s. The fact that Commission staff are at the forefront in extending GTAP to analyse GATS-related policy issues demonstrates the PC's commitment to, and capability of, remaining at the frontier of applied empirical policy analysis.

The extended model has been given the name FTAP, the F referring to its inclusion of foreign direct investment. Specifically, the additional features include increasing returns to scale, imperfect competition, capital accumulation, international borrowing and lending, and the ownership by country of the economic activity in each country/country group.

With those additional features comes the requirement to somehow obtain (dream up?) many more parameter values. This adds to the already major problem of not having a comprehensive database of bilateral services trade and investment flows for the GTAP model, although that problem is expected to be reduced in the next version of the database (Version 5 — see McDougall 2000). Parameter values in FTAP are simply assumed for such items as elasticities of substitution in

consumption between domestic and foreign products (constant value of 5 used for all goods and services), between foreign locations of production (assumed to be 10 for each location), between ownership categories in a particular location (also assumed to be 10), and between individual firms of a particular ownership and location (assumed elasticity of 15). Likewise on the supply side, there are assumed elasticities for investors who divide their wealth between physical assets (including natural resources) and financial assets (bonds). That requires selecting parameter values to reflect how investors choose the industry in which to invest, whether to invest in that industry domestically or abroad and, if abroad, which overseas country in which to invest. For simplicity the authors have set all of those semi-elasticities of transformation on the supply side at between 1 and 1.4.

How important are each of those assumptions to the model's results? And what about the assumption that capital is less than perfectly mobile between sectors within an economy, even though sector-specific capital is internationally mobile between economies? That assumption is sensible for at least some sectors, being consistent with the Markusen theory of knowledge capital giving rise to multinational corporations, but how important is it? Systematic sensitivity analysis is clearly called for.

D15.2 Measuring and modelling the barriers to trade in services

To all these data and parameter challenges need to be added the challenge of finding estimates of barriers to services trade (the second part of the paper). The authors draw on estimates for banking and telecommunications from Findlay and Warren (2000) and assume them to be representative for all tradable services. They further assume the rest of services and electricity, water, gas, construction and housing are nontradable internationally. In table 15.1 they tabulate those barriers to services trade and investment and distinguish barriers to establishment (taxes on capital, domestic versus foreign) from barriers to ongoing operation (taxes on output of local firms, whether domestic or foreign owned, of the same size for both domestic and export sales). This way of including the measures is designed to be consistent with the four modes of service delivery identified under GATS. For symmetry — even though it is not relevant to WTO commitments at this stage — economists would also like to see foreign investment in goods included in the model as well, not least to get an idea of the relative economic importance of barriers across different forms of cross-border flows of goods, services and capital.

What turns out to be a crucial assumption in this specification of measures is the treatment of taxes on service exports. Those tax revenues are assumed to stay as rents in the country imposing the tax. An alternative assumption is that costs of production are raised by that amount, in which case there are no rents to be

transferred. The truth is probably somewhere between those two extreme assumptions so, until we know more about the true nature of these distortions, it might be wise to report welfare results under both assumptions (see below).

What do these assumptions imply about the size or extent of distortions to prices of services, as distinct from goods, post-Uruguay Round? If the banking and telecommunications barriers as measured are indeed representative of all services, they suggest services price distortions are at least as large as those for goods in most countries, and are larger for exporters in developing countries than those in advanced economies.

D15.3 Liberalising barriers to trade in services

Turning to the third part of the paper, what can we learn from the empirical results about the relative welfare costs of services versus goods trade distortions? The welfare significance of the price wedges depend on several things: the relative size of the sectors, the extent to which its outputs are inputs to other sectors, and the way in which the price wedges shown in table 15.1 are generated. If as assumed those wedges are generated like an export tax, then the welfare losses are confined to the equivalent of Harberger triangles in the standard supply/demand diagram. If, however, they were assumed to be caused by an avoidable cost burden (as, for example, in Hertel 1999), then reform could also generate technical efficiency gains. Since the latter could be sizable relative to the former, the current study's global results should be considered as possibly very conservative lower-bound estimates.

Furthermore, the welfare estimates depend on whether it is assumed the rent is retained by incumbent firms, is appropriated by the government via taxation, or is passed from one country to another via transfer pricing by multinational firms. The present study assumes rents on outputs stay with the country of location while after-tax rents on capital go to the country of ownership. This set of assumptions has important implications for the international *distribution* of the global gains from reform.

What are those estimated effects of removing the services trade and investment distortions that the authors have included in their FTAP model? They are not obvious a priori because they depend, among other things, on the share of services in each country's GDP, the share of each country in global services output, the extent to which services distortions discriminate against foreign suppliers (since not all measures so discriminate), and the extent to which services at home and abroad are inputs into production of other goods or services.

The FTAP database, as summarised in figure 15.1, suggests much of the impost will be on the service sector itself, as that sector is a relatively intense user of services inputs. This contrasts with goods trade liberalisation, where one sector's reform usually hurts that sector, while helping other sectors. For politicians, that difference between goods and services could be very important — there may be less reluctance to liberalise services if it means any displaced workers can move to another service industry in a nearby location. Recall, however, that a multilateral liberalisation of services markets would also have a terms of trade effect much bigger than under a unilateral reform. Decomposition analysis would be helpful in indicating the relative contribution of that effect.

Decomposition analysis also would be helpful in indicating the relative contribution of measures affecting different sub-sectors of services. Some preliminary analysis by Francois and Wooten (1999) suggests, for example, that maritime transport services reform, through lowering ocean transport costs (by breaking the current cartel-type arrangements), could give a substantial boost to goods trade and to economic welfare of developing countries (since most of the ships are owned by rich-country firms).

The present paper tabulates primarily sectoral output results. (See Hanslow (2000) and Verikios and Zhang (2000) for more detailed early results.) But it also mentions a crucial welfare result, namely, that the United States (and the European Union in the full model results) are projected to be worse off under services trade liberalisation. Given the keenness of the United States and European Union to see further services reform, this seems unlikely to be the case in practice. Presumably, the result stems from the authors' assumption that wedges between domestic and international prices are due to quantitative restrictions that can be modelled as export taxes. An alternative extreme assumption, that the wedges are due to cost-raising inefficiencies, would not give rise to the large rents that are generated in the current specification of FTAP and so would not be as likely to generate losses from reform for large traders of services.

This unfortunate welfare result should not detract from the worth of this paper, however. On the contrary, it simply underscores the authors' point that this modelling effort is still in an early stage of development. We are indebted to the team working on this project for exposing the much greater complexity of the task of modelling services and capital flows as compared with just goods trade flows. Their study throws out a challenge to numerous colleagues: to statisticians, to improve bilateral foreign investment stock and flow data by sub-sector; to trade theorists, to improve our theoretical models of world trade by examining how standard two-sector results change with the inclusion of services trade and investment; and to econometricians, to provide better estimates of elasticities of

substitution and transformation. Prioritising that theoretical and econometric work would be helped by a taxonomic paper by CGE modellers on the sensitivity of results to changes in their assumed parameter values.

References

- Findlay, C. and Warren, T. (eds) 2000, *Impediments to Trade in Services: Measurement and Policy Implications*, Routledge, London and New York (forthcoming).
- Francois, J.F. and Wooten, I. 1999, Trade in international transport services: the role of competition, Paper presented at the First Annual Conference of the European Trade Study Group, Erasmus University, Rotterdam, 24–26 September (www.etsg.org/etsg/detailedprogram.htm).
- Hanslow, K. 2000, The structure of the FTAP model, Paper presented at the Third Annual Conference on Global Economic Analysis, Monash University, 28–30 June.
- McDougall, R. 2000, International services trade data for CGE modellers, or, entropy methods for data reduction, Plenary paper at the Third Annual Conference on Global Economic Analysis, Monash University, 28–30 June.
- Verikois, G. and Zhang, X.G. 2000, Structural impacts of liberalising trade in services, Paper presented at the Third Annual Conference on Global Economic Analysis, Monash University, 28–30 June.

Discussant — *Will Martin*

World Bank

This is a really interesting and insightful paper. In the limited time available, I would like to raise some questions on the modelling aspects of the paper.

D15.4 Model parameters

An important question is how to get estimates of the parameters that the model uses. Econometric estimates of many of these parameters are unlikely to be available in the foreseeable future. Maximum entropy techniques (eg Liu, Arndt and Hertel 2000) may provide a means to obtain some of these parameters from actual data. But it is a real concern when there are no estimates available for many key parameters.

Given these uncertainties associated with many key parameters, there is a need for sensitivity analysis. I'm a little uncomfortable with the relatively wide range used for the substitution elasticities — between 5 and 10, especially given the narrowness of the range and apparent precision of the transformation elasticities (1 and 1.4). Are these parameters so different? This is where sensitivity analysis comes in — how sensitive are the results to these critical parameters?

D15.5 Modelling trade barriers

The specification of wedges is very interesting. It incorporates different sets of instruments: taxes on capital, taxes on output, and discrimination between domestic and foreign firms in either tax. One question, however, is whether restrictions on the movement of natural persons — the important Mode 4 of GATS — can be modelled adequately in the current framework.

Another question is why the model does not have any black hole rectangle losses, especially given their prevalence in the paper by Tony Warren and Christopher Findlay (see chapter 14 in this conference volume). Incorporating black hole rectangles of economic loss will have implications for both data collection and modelling efforts. There are likely to be rectangle losses or efficiency losses resulting from many services regulations.

There is a great need for further work — especially in providing new estimates of parameters and trade barriers. There is also a need to distinguish between rent dissipation and taxes — elements critical in modelling service trade liberalisation.

Reference

- Hertel, T. 1999, Potential gains from reducing trade barriers in manufacturing, services and agriculture, Paper presented at the 24th Annual Economic Policy Conference, Federal Reserve Bank of St. Louis, 21–22 October.
- Liu, J., Arndt, C. and Hertel, T. 2000, Estimating trade elasticities for GTAP: a maximum entropy approach, Paper presented at the Third Annual conference in Global Economic Analysis, Melbourne, 26–29 June.

General discussion

The general discussion focused on:

- the model parameters and sensitivity analysis;
- the modes of service delivery; and
- the modelling of trade barriers.

The model parameters and sensitivity analysis

Participants recognised the need to obtain estimates of elasticities and conduct sensitivity analysis of model parameters to ensure that estimates of substitution elasticities were not out of line. The modelling team noted that the elasticities of transformation were more heavily nested than the substitution elasticities and appeared reasonable.

One participant also noted that it was important to have different substitution elasticities for goods and services, and for individual service sectors. Services were different from goods in requiring face-to-face contact, and this should be relevant. The modelling team noted that it was a deliberate intermediate research strategy to begin with a simple model structure having uniform estimates of elasticities across goods and services.

In relation to sensitivity analysis, the modelling team suggested that it would be easier to conduct sensitivity analysis with respect to model parameters than it would be with respect to data on trade barriers. In this respect, there was a more pressing call for estimates of barriers than estimates of model parameters.

The modes of service delivery

One participant noted the need to provide separate treatment of the different modes of service delivery. While commercial presence was captured separately, other modes of service supply were combined into the category of ongoing operations. However, different modes of supply faced different trading regimes. Cross-border delivery of services, where it was feasible, was typically very liberal. For example, data processing and software development faced no restrictions in cross-border supply. In contrast, the movement of natural persons — which was crucial in areas such as construction services and health services — was a highly restricted mode of supply. To anticipate how countries' patterns of trade would change according to comparative advantage, some of the predictions about developing countries expanding their share would depend on the liberalisation of the movement of

natural persons (mode 4). Therefore, it was valuable to separate the different modes of service delivery according to their feasibility and restrictiveness.

In reference to Will Martin's comment on restrictions on labour movement, the modelling team clarified that the wedges were currently included with the taxes on capital. This treatment reflected the idea that restrictions on the permanent movement of people also prevented the movement of capital. Foreign firms could be prevented from establishing in a market if they faced restrictions on bringing in their own management and personnel. A separate treatment of these restrictions was needed, however.

The modelling of trade barriers

On whether barriers to trade in services should be modelled as rent-creating or cost-creating, the modelling team suggested that the econometric studies that quantified the effects of barriers should indicate whether the restrictions were rent-creating or cost-increasing in their first round impacts. The study on banking services (Kalirajan et al. 2000) explicitly examined the impact of trade barriers on banks' net interest margins and showed directly that those margins were being inflated by restrictive regimes. A study on aviation services (Johnson et al. 2000) had shown that trade restrictions were at least partially cost-increasing. Recent studies on wholesale and retail trade (Kalirajan 2000) and the professions (Nguyen-Hong 2000) were also suggestive that some restrictions in these sectors were cost-increasing rather than rent-creating. Other empirical work in the pipeline would seek to estimate such effects separately.

In reference to Kym Anderson's comment on the need to incorporate foreign investment in goods into the model, the modelling team clarified that the model incorporated foreign direct investment flows on goods trade, but estimates of the investment distortions were not yet available.

In reference to Kym Anderson's question of whether banking and telecommunications barriers were representative of barriers to trade in services, the modelling team noted that work was under way to separate out various service sectors and to assign each sector a measure of its own individual barriers.

References

Kalirajan, K. 2000, *Restrictions on Trade in Distribution Services*, Productivity Commission Staff Research Paper, AusInfo, Canberra.

—, McGuire, G., Nguyen-Hong, D. and Schuele, M. 2000, ‘The price impact of restrictions on banking services’, Chapter 13 in Findlay, C. and Warren, T. (eds), *Impediments to Trade in Services: Measurement and Policy Implications*, Routledge, London and New York (forthcoming).

Johnson, M., Gregan, T., Gentle, G. and Belin, P. 2000, ‘Modelling the benefits of increasing competition in international air services’, Chapter 8 in Findlay, C. and Warren, T. (eds), *Impediments to Trade in Services: Measurement and Policy Implications*, Routledge, London and New York (forthcoming).

Nguyen-Hong, D. 2000, *Restrictions on Trade in Professional Services*, Productivity Commission Staff Research Paper, AusInfo, Canberra.

PART G

ACHIEVING BETTER
REGULATION OF SERVICES

16 Approaches to improving regulation: regulatory competition or regulatory harmonisation? A forum

The final session of the conference consisted of a forum whereby key participants raised issues that they thought would contribute to achieving better regulation of services. The session concluded with a general discussion.

Christopher Findlay

Australian National University

This conference has highlighted several themes and issues in regulation of services and possible approaches to reform. I will briefly review the major themes that participants thought to be important issues that are not yet resolved.

Themes

This conference has stressed the value of a systematic approach to reviewing regulation. Paul Coghlan's paper (chapter 3) provided participants with confidence about the ability to do this and the methodology involved.

A key factor arising from the studies of regulatory arrangements in particular service sectors was the capacity of sectoral interests to organise and get their agenda in place. That process was clearly at work in some service sectors, such as retail pharmacy (chapter 9).

The impact of technological changes was also recognised. New technology affects the way service transactions are conducted and opens up new ways of doing business. But new technology also has the capacity to unravel and circumvent many existing policies, making it more difficult to apply standard policy instruments to achieve policy objectives. This is forcing people to think again about how to regulate.

There were some nice examples of the contrary effects of bad policy, where the policy ended up achieving the opposite of the desired outcome. The paper by Patrick Messerlin (chapter 12) showed that interventions to protect French culture

had the unintended effect of making the French film industry more like Hollywood. This could be a common outcome when the attempt was to achieve many targets with a single instrument, or where insufficient attention was given to the incentives being created.

A common story about the service sector is the way services are delivered, the importance of contact between producers and consumers and, therefore, the kinds of policies that are relevant to manage or control trade and investment in services. Regulation is an important part of that portfolio. Participants regularly asked whether, and how, services differed from goods.

Finally, while it was recognised that regulatory reform was clearly a good idea, especially in some areas, there was an understanding of the value in using international instruments to help manage the regulatory reform process — to get it going, to direct it, and to lay down some principles to achieve progress.

Outstanding issues

The themes of this conference also bring out some issues that are not resolved.

There is clearly further scope for trade in services to grow because of new developments in technology. There is a growing Australian constituency of service providers with an interest in market access offshore, and increasing numbers of people overseas interested in getting access to Australia.

The globalisation of enterprises in the service sector and the opportunities presented by new technologies will put increasing pressure on policy and its implications for trade and investment. There is scope for dispute between Australia and our trading partners to resolve these matters. It seemed to me that this trade perspective remained unexplored in a number of papers on services. The focus was still mainly on domestic markets and how they worked, but lurking in the background is an important trade agenda that needs to be considered.

With respect to trade policy disputes, in principle there is a framework available for resolving them. However, in thinking about how the framework might work, the tests or principles are not really well resolved. To what extent should regulations be left to domestic policy processes and reviews, or are there benefits from having international rules or horizontal principles? This requires further discussion.

Ultimately, achieving better regulation of services will require both well-founded international rules and rigorous domestic policy. I think that better regulation has to come from above, from the multilateral system, the articles of GATS, the reference

papers and other international instruments. Better regulation also has to come from below, from the way rigorous domestic evaluations are undertaken.

Alan Moran

Institute of Public Affairs

A theme of this conference was the recognition of the costs of regulation and the role of market mechanisms in efficient service delivery. The papers by Paul Coghlan (chapter 3) and Neil Byron (chapter 8) tell us that the costs of regulation are more likely to outweigh the benefits. The difficulty in trying to dismantle regulatory privileges was also highlighted in the case of retail pharmacy (chapter 9).

I would like to make three observations.

First, there is no consensus between what I might call the *competition enforcers* and the *Austrians*. The competition enforcers seek to promote competition as a goal in itself and to intervene to promote competition in the market. For example, competition enforcers would seek to promote competition in essential facilities with the goal of forcing prices down and squeezing out rents. In contrast, the Austrians are much more concerned about allowing entrepreneurship to seek out rents, to foster investment and promote better productivity.

Second, arguments based on supposed externalities or market failures are still being enlisted to justify the regulation of services. The Gambling Report by the Productivity Commission (chapter 11) and the papers on culture by Patrick Messerlin (chapter 12) and Pierre Sauvé (chapter 13) all make this type of case.

Third, it needs to be recognised that regulatory reform is very much on the back foot in some areas. The whole notion of economic rationalism has been pilloried. In New Zealand, reform is being wound back. There is a danger of seeing that happening in Australia. Australia certainly has a pause in reform at the present time. To go forward, we have to find ways to make reform attractive again and create the political support.

Kym Anderson

University of Adelaide

This conference has highlighted issues at the national and multilateral level. The work of the Productivity Commission and its predecessors, through its inquiry process and through the activities of the Office of Regulation Review, have had an enormous impact in getting business and other stakeholders in the service sector to think more like economists about the current regulatory arrangements.

This was evident when the Industries Assistance Commission (the predecessor of the Productivity Commission) first formed in 1973, with both manufacturing and agricultural stakeholders becoming much more articulate about trade reform issues. This had an effect on the direction of liberalisation.

There are parallel developments at the multilateral level. Patrick Messerlin's paper compared culture with agriculture (chapter 12), showing how large filmmakers benefit from the protection of audiovisuals in France, in the same way that large farmers benefit from the Common Agricultural Policy. Then Pierre Sauvé (chapter 13) gave a defence of protecting culture reminiscent of the multi-functionality debate within agriculture, where this is a trite justification for protecting farmers.

More generally, an important issue is whether the WTO should be involved in domestic support of the service sector. This debate also goes on with respect to goods and agriculture. Under the WTO, some disciplines were developed in the Uruguay Round Agreement on Agriculture. The experience with the multi-functionality debate suggests that the more attempts there are to discipline domestic supports within the WTO, the more groups will find ways around those rules and provide support in an even more inefficient way.

I agree with Richard Snape's comment in the last session, that it may be best to leave domestic support out of the WTO and focus on issues at the border. When the views of economic rationalists are pressed on the domestic community, more articulate voices emerge against these views, and the policy outcomes are moot.

Will Martin

World Bank

An important issue that needs to receive more attention is the form of intervention or regulation designed to achieve policy objectives.

If the need arises for regulation to protect consumers, it is important that such a regulation provides incentives to directly encourage the protection of consumer interests. For example, in the taxicab industry, the policy could aim to create some sort of bond that gives the owner of the cab the incentive not to engage in practices harmful to consumers. But the way regulations have been enacted in many cases is to regulate the quantity of taxi licences. This then creates a stake in the system for the taxi owner, while not doing much for the driver. The quantitative restrictions simply create enormous transfers.

If regulation is of the wrong form, it is like using a quantitative restriction rather than a tariff. Therefore, it is very important to think about the specific form of intervention right at the beginning.

Richard Snape

Productivity Commission

Developments in trade policy around the world and in Australia over this century provide important lessons for adopting appropriate approaches to regulatory reform of services.

History of trade policy and lessons

Why was regulation in trade so bad after the Second World War? It was partly the legacy of the two world wars, since industries had been built up behind high trade barriers during those Wars. Monetary crises also played a part, since monetary stability is an indispensable condition for good trade policy. In the 1930s, attempts to generate employment through policies such as competitive devaluation also contributed to the problem.

In the early years after Second World War, many misguided economic and political policies continued. Those years saw a great deal of Marxist ideology and inward-looking economic development policies. These factors contributed to flawed policies for trade and development in developing countries and rampant protectionist policies.

How has it all changed? One reason is post-war prosperity. In this context, it is interesting that trade policy has not been set back by the recent Asian crises. The reasons for this may provide some lessons.

The period after the Second World War also saw a high degree of regulatory cooperation, which had not existed before. Regulatory cooperation took place through the GATT on trade policies and the International Monetary Fund (IMF) on monetary policies. In the 1950s, the formation of the European Community was a major factor in breaking down trade barriers within Europe, and contributed to a great improvement in regulatory policies on trade, at least outside agriculture.

Improved economic understanding had a major effect on trade policy. There was a greater understanding and appreciation of economics and a great deal of quantification of the effects of government policies.

Good regulatory practices provide demonstration effects for other countries and are an important facet of regulatory competition. Developing countries with outward-looking policies achieved economic success and the lessons convinced many other countries to adopt liberal policies. In fact, the economies with outward-looking trade policies triumphed and their policies therefore triumphed over the earlier inward-looking development philosophies.

Leadership has been very important to the success of the multilateral trading system. Indeed, the recent collapse in Seattle largely reflects an absence of leadership by the major democracies. The United States played an important role in the post-war period in moving towards better regulation in its trade policies. There was a relatively enlightened attitude in the United States to trade policy issues, although this enlightenment sometimes waxed and waned.

A significant development in US trade policy was the passage of the reciprocal trade legislation in 1934. The legislation essentially removed trade policy from Congress and placed it under the responsibility of the Administration. This had an important consequence. Trade policy was no longer considered as a tariff-by-tariff issue, as it had been previously conducted in Congress. Placing trade policy under the responsibility of the Administration allowed trade policy to be considered in a broad arena, beyond specific issues and interests. Of course, the industrial strength and prosperity of the United States was also important in the success of its trade policy.

A broad focus

One lesson is that good regulation in trade policy has been most easily achieved when trade policy has been considered in a broad arena and not issue-by-issue, or product-by-product. For example, liberal trade within Europe was achieved when trade policy was considered as part of the Treaty of Rome and subsequently as part of the wider European Community Agreement. Moving away from specific issues into the broad arena helps to promote good regulatory practice.

In the GATT, trade policies have been adopted in the form of general rules. The GATT was a general agreement that essentially sets rules for most countries and products. The Agreement was successful because the general rules provided opportunities for sectoral tradeoffs and other advantages. This is an additional reason to move away from specific issues into a general framework.

Transparency

In Australia, there was a major improvement in economic understanding of the effects of protectionist trade policy in the postwar period. Institutions such as the Tariff Board and its successors have had important educative functions and Australian economists, such as Max Corden, also contributed to better economic understanding of government policies.

Public exposure and transparency of policies are therefore important. Policy transparency helps to inform debate and mobilises opposition to protected interests. For Australia, liberalisation in other countries has also had demonstration effects. So competition in regulatory policies can also have an educative role.

Regulatory coordination

Coordination is also important. Regulatory coordination is best achieved when it is placed in a wider arena that goes beyond particular interests. One problem of the GATS is that while it contains general rules, those rules are very weak compared with those of the GATT, so that the GATS leads to industry-specific negotiations that inhibit sectoral tradeoffs. The focus on a single issue or industry greatly inhibits effective negotiation and coordination.

Therefore, the framework for international coordination of regulation must be designed with great care. It needs to have a wide arena and not be narrowly focused. One lesson from air services is that because the air service agreements are narrowly defined, it is impossible to obtain tradeoffs against other industries. Once an industry-specific approach has been locked in, it is difficult to move away from the narrow framework to achieve regulatory reform.

Regulatory competition

It is often argued that competition between jurisdictions will lead to migration of people to countries where there is good regulation. However, there are only a few issues that, on their own, would cause people to migrate. People may migrate because of a whole regulatory framework, as in the case of centrally planned economies. But there may not be great benefits in relying on the migration of people and the benefits of competition to reform specific regulatory structures.

Business will migrate rather more easily from one country to another with regulatory competition, but that does not always generate the best form of regulation. Business may migrate to avoid regulation, even though that regulation might have been in the general interests of the community.

Thus, there are two caveats to the benefits of competition between regulatory systems. Business migration may not always generate the best system, and there are very few single issues that will cause people to migrate.

David Hawes

QANTAS

This conference has brought back to me various memories of negotiating the GATS and the debates held at the time. I detect an interest on the part of some people to try rewriting the history of the GATS negotiation. I would ask you not to do that. However, I would encourage efforts to complete the outstanding provisions and review and revise parts of the text, because I would be among the first to acknowledge that the GATS is not a perfect instrument.

General versus sectoral approach

The weakness of the initial liberalisation process in the GATS reflects a number of considerations. The players involved from 1986-93 were on a voyage of discovery. Many countries had highly complex regulatory frameworks for services, while others had no regulations in place. The aim, through largely consensus-based negotiation, was to persuade these various players to:

- sign on to an agreement, the outlines of which they could barely see for many years; and
- give commitments that would liberalise, but would also give stability and predictability to their expectations as signatories to the WTO agreements.

On the question of having general or specific commitments, my experience with aviation suggests there is a need for a sector-specific arrangement to bring the players from the arcane and anachronistic 1940s system of regulation towards a free and open system, more akin to that applied in other sectors.

How can this transition be made? It cannot be done quickly. The arrangement needs to provide the air transport sector with confidence to come on board. The two most pressing concerns of the sector are the most favoured nation (MFN) principle and the prospect of being bound up in cross-sectoral tradeoffs with other sectors. The airline sector has had generations of people working airline to airline, and aviation bureaucrat to aviation bureaucrat. These broader notions have been major concerns.

A transitional arrangement based on conditional MFN or an open club arrangement (which could grow to include new members) has been suggested as an option which might eventually provide the confidence for moves towards an over-arching

multilateral agreement in the WTO or some other forum. While previous efforts to create a multilateral international transport agreement failed, the world is changing rapidly, the airlines are being required to adapt to globalisation, and some are under threat.

Harmonisation of competition policy

The airline industry in Australia is required to compete in an increasingly competitive international environment, while subject to a competition policy that is domestically driven, democratically rooted. While Australian competition policy aims to protect the Australian consumer, some of our major international competitors are subject to different forms of competition policy or are headquartered in markets where such policies do not even exist. This is a major problem. Increasingly, we will hear calls for competition policies to be harmonised internationally in order to give the industry confidence to go hand in hand with the liberalisation process.

Pierre Sauvé

Harvard University

Fifteen years or so into the GATS, I think that we are still very much in learning mode. Looking back on the negotiating history of services liberalisation in the GATS and at a regional level in NAFTA, little was known of the contribution that some regulation could make in underpinning and complementing trade and investment liberalisation. The conference also shows that we are still very much learning by doing. First movers on privatisation (eg the United Kingdom) are now experiencing first mover disadvantages and realising the mistakes made in the design of institutions that would support privatisation. This is also being realised in the trade context.

Aaditya Mattoo has written extensively about how the liberalisation of investment and access regimes has encouraged changes in ownership, but paid insufficient attention to conditions of contestability and barriers to entry. Privatisation has too often shifted rents from the public to the private sector, without promoting efficiency gains. So we are still in learning mode and we should not be surprised at this.

Another message of the conference is the need for additional work and research on issues of regulatory reform of services. The OECD is in a very good position to undertake this type of work in a comparative setting and I hope to enlist your help in making sure that discussion can take place in an international setting. Clearly this could also be done in the WTO, but the OECD seems a very good place to start.

General versus sectoral approach

On the debate between sectoral and generic approaches, I am very sceptical about the scope for generic approaches to regulatory practices in a WTO setting at this point in time. I think we have to be patient and rely predominantly on sector-specific experimentation.

This is because in most countries there is still a need to promote what may be called a ‘culture of regulatory buying’. Trade negotiators and regulatory authorities still need to learn to speak a common language. In most countries, this does not happen. So excessively generic approaches anchored in GATS are likely to increase resistance on the part of the regulatory communities to reforms that facilitate market access. This observation reflects my own experience of the regulatory resistance in Canada to trade and investment liberalisation mandated by international rules or agreements.

The resistance on the part of the regulatory communities should not be underestimated. Thus, if we can experiment with sector-specific approaches in a number of sectors, as was done in the telecommunication sector, a degree of similarity in regulatory approach could be achieved, that could perhaps then be turned into generic disciplines. However, we should anticipate emphasis on sector-specific work in the first instance.

Regulatory objectives

As Paul Coghlan’s paper (chapter 3) suggested, regulatory objectives are often unclear. In NAFTA, a measure to promote transparency was that whenever member countries scheduled a reservation or maintained measures that deviated from the free trade principles of national treatment or market access, they were required to reveal the regulatory objective being pursued.

While the approach seems straightforward, it was enormously difficult to agree on the categories of regulatory objectives being pursued. In the WTO context, one option is to draw up an illustrative list of regulatory objectives in the same way as an illustrative list of national treatment restrictions is being considered. This would enable countries to speak a common language and to understand the commonalities and, in some instances the differences, in regulatory approaches and objectives. Some useful work remains to be done in this area.

Regional experimentation

Considerable scope exists for transferring best practice in regulation, and in the interface between regulation and trade, from the regional to the multilateral level. We can ‘let a thousand regional flowers bloom’. But we are still in learning mode, and the experimentation in different regional settings, such as APEC, NAFTA and the EU, can inform how some of the challenges can be resolved in a multilateral setting.

This can be done without being unduly concerned about the MFN consequences, partly because the multilateral system does not impose very onerous disciplines on regional experimentation. Any attempt to promote domestic regulatory conduct that facilitates market access is likely to be WTO compatible. But there is a need to experiment and to take advantage of the opportunity to do so in smaller settings, because such experimentation is easier in smaller settings.

Necessity and needs tests

A debate on the necessity test in Geneva cannot be avoided. There is no way to make progress on liberalising regulatory barriers and settling disputes arising from the onerous nature of those barriers, without agreeing on the criteria to determine what is necessary regulation. A necessity test will not question regulatory objectives per se, but criteria are needed to determine when a measure is ‘necessary’. This is a longer term objective.

Similarly, there is a need to consider economic needs tests at the multilateral level. Insufficient attention is being paid to barriers that are often non-discriminatory, but nonetheless have potentially significant restrictive effects on competition and markets.

Harmonisation versus mutual recognition

My experience suggests that harmonisation is a great objective to promote when countries are not serious about liberalisation. A comparison of the liberalisation experience in telecommunications and land transport within NAFTA illustrates this difficulty.

In telecommunications, it was recognised that liberalising telecommunications markets cannot be achieved without paying attention to the way networks interconnect. The solution was to seek mutual recognition of the certification, testing and licensing procedures without which the systems would not interconnect. That process was achieved in a short timeframe of two years.

In land transport, where there was strong resistance to liberalisation from the Teamsters Union, the NAFTA countries opted for harmonisation of road safety standards. Eight years after completing the agreement, a common regime of road safety rules has yet to be developed.

If NAFTA countries were serious about liberalising land transport, they would have opted for a system of mutual recognition — a difficult enough task when countries are at different levels of development. But aiming for harmonisation made the task almost impossible.

Transparency

Transparency is an important mechanism to facilitate liberalisation of regulatory barriers. In NAFTA, there was a negative list approach to liberalisation. Countries were required to document all measures that did not fulfil national treatment, market access or other obligations of the Agreement. A failure to list non-conforming measures meant that a country was automatically bound by the general terms of the Agreement.

This approach to liberalisation forced a degree of dialogue between the trade negotiating community and the regulatory community that had never occurred before in Canada. It was a healthy dialogue about the optimality of domestic regulatory regimes. When there was significant evidence of non-conforming measures, regulators were asked simple questions — when was the measure last used? what was the objective? and are there alternative ways of achieving the objective? The process was a very useful exchange of views that helped the participants to speak the same language.

In the GATS, a shift from the current hybrid approach to a negative list approach is unlikely to happen. That said, Australia, Canada and those countries that take transparency seriously should push for an ‘operation transparency’ — that is, the production in each GATS member country of a non-binding list of non-conforming measures. This list would form a basis for developing formula-based approaches to liberalisation in the future. It would also provide a means to compare respective approaches to regulation, so that where there was considerable commonality, then common ways of lessening restrictive effects of these regulations could be devised.

These approaches to liberalisation are not regulatory neutral, nor are they trade neutral. Finding ways of documenting the kinds of practices that are maintained in our respective economies would go a long way towards the efficient regulation of services.

Patrick Messerlin

Institut d'Etudes Politiques de Paris

Mutual recognition agreements

Mutual recognition is not followed per se in the European Community. There is always a tradeoff with other Community goals, such as promoting the Common Market, safety, small enterprises, etc. Thus, there is always something that dampens the effects of mutual recognition agreements.

There are two layers of mutual recognition agreement. The first layer involves designing new norms and new regulations. The second layer is mutual recognition on testing, certification and implementation of the new norms. To date, the mutual recognition agreements at the WTO level are essentially of the second type, about testing and certification. It may be inevitable that the WTO will go back to the first layer of designing new norms and regulations. But while the three papers that the EC prepared for the Seattle meeting focused on the need for domestic disciplines, they did not mention mutual recognition.

Degrees of mutual recognition

There are three layers of mutual recognition. The first is harmonisation, which applies to some goods. Harmonisation is a very long process, unlikely to be pursued further in the WTO. One interesting consequence of following the harmonisation route, at least in the European Community, is that it is not possible to switch to other arrangements.

The second degree of mutual recognition is what I call regulatory convergence and the third is regulatory competition.

The difference between regulatory convergence and regulatory competition is the extent to which the agreement incorporates common provisions. Regulatory convergence brings about a substantial, common core of provisions. Regulatory competition does not generate new norms or standards — it essentially leads to mutual recognition of certification and testing.

It seems to be a law that negotiation always produces regulatory convergence and never stops at regulatory competition, at least in the history of the European Community. And in negotiating about provisions, there is a strong tendency to focus narrowly on one sector and to forget about all others. The result in Europe has been chaotic common core provisions that are far from ideal. The EC then has to

ask the Court of Justice to differentiate among the various provisions in the initial text of the agreement, and re-negotiation starts. It is an extremely painful process.

My view is that where there has been attempts at regulatory convergence, there has been no liberalisation except through technical progress. Thus there has been liberalisation in telecommunications and associated industries, but very little progress in the other sectors.

European Community

The Treaty of Rome has a focus on so-called infrastructure services, such as telecommunications, banking, airlines, land transportation and audiovisuals. As a result, when a new infrastructure service such as e-commerce develops, there is a problem in integrating the new sector into the existing regulatory approach. For example, now that films can be downloaded from the internet, the issue is whether this constitutes audiovisual services, e-commerce services, telecommunications services, or whether it is a good, such as a substitute to a cassette. The EC has taken the safest route by declaring that this is a service, which means that the least liberal regulatory regime applies.

Finally, Sanitary and Phytosanitary (SPS) disputes within the WTO should be kept in perspective by noting the even bigger dispute within the EC relating to mad cow disease. As a result of this, the EC has a total ban on meat exports among Member states. I think perspective is also important in terms of mutual recognition agreements.

Aaditya Mattoo

World Bank

1 Approaches to eliminating regulatory barriers

Consider harmonisation, mutual recognition, and the necessity test in the multilateral context, which is similar to the so-called 'policed regulation' in the European context.

Harmonisation

Harmonisation is a long way away, as Pierre Sauvé said. Where international standards exist, for instance in banking or maritime services, they are seen more as a necessary first step, rather than as a sufficient condition for market access. It is only in accountancy services that some progress has been made. But at the level of

the WTO, the best that can be done is to create a presumption in favour of the choice of such standards, and this has happened in the Technical Barriers to Trade agreement (TBT) and the SPS agreement. A weaker approach is the GATS provision that stipulates that, in determining whether a country's standards are excessively burdensome, account shall be taken of whether or not the country has met the international standard.

Mutual recognition

I was a little surprised by Patrick Messerlin's point that the EC has not brought the issue of mutual recognition into the WTO context. Mutual recognition cannot be made to happen. Indeed, it depends largely on the incentives of the individual countries to conclude mutual recognition agreements. For instance, it is quite likely that more mutual recognition will occur in the future, for example, between the European Union and the United States in areas such as accountancy services, to encourage gentle intra-industry trade. However, the incentives to conclude a mutual recognition agreement between countries such as the European Union and the Philippines or India are likely to be extremely weak, even when there are no significant differences in standards, because the adjustment costs from inter-industry trade are likely to be greater. Thus, what is likely to happen is trade according to a pattern of mutual trust rather than a pattern of comparative advantage.

Mutual recognition agreements are like sector-specific preferential arrangements, with the same trade-creating and trade-diverting effects. But preferential tariff arrangements in goods displace high-tariff imports from third countries, so the resulting loss of tariff revenue disciplines their use. Mutual recognition agreements do not involve the elimination of revenue-generating barriers, so there is not the discipline of lost revenue. Nevertheless, preferential arrangements in services can still have trade diversion costs by excluding supply from third countries.

This is why the existing WTO rules on mutual recognition agreements are very important. The GATS provision on recognition (Article VII) strikes a balance in a way that the TBT agreement does not. Article VII is permissive by saying that countries are free to conclude mutual recognition agreements because of the trade-creating effects. On the other hand, it requires that those countries must not use the agreements as a form of discrimination and they must allow third countries access to such agreements. In this respect, Article VII of the GATS has a very desirable feature of openness towards third countries that Article V on regional integration does not. Article VII is more like open regionalism, which contrasts with the closed group nature of regional integration agreements.

Interestingly, many countries, including Australia and New Zealand, are choosing to notify their mutual recognition agreements under Article V on regional integration agreements, rather than under Article VII. An important question is whether this means that those countries escape the disciplines of Article VII. This is potentially an important question, because it determines whether the interests of third countries are sacrificed.

2 International cooperation on competition policy

The GATS has so far treated regulation as a handmaiden of market access. In the terminology of the Warren and Findlay paper (chapter 14), the sins of omission — the failure to have competition policy or access rules — are really about protecting market access rights, and the sins of commission are excessive barriers that undermine market access.

However, where there is the greatest need for international cooperation is where there is an exemption from national competition law for international cartels, such as in maritime transport.

Some recent work suggests that while government restrictions adversely affect maritime transport costs, private anticompetitive practices have an even stronger effect (Fink, Mattoo and Neagu 2000). A large country like Australia can remedy this through its national competition laws. However, a small country like St Lucia is at the mercy of international cartels. Such spillover problems provide the strongest case for multilateral disciplines, because much of the other regulatory reform can be accomplished at the national level. Article IX of the GATS, which deals with anticompetitive business practices, is extremely weak. This outcome is largely because of the opposition of countries like the United States and the EU, which exempt these cartels from the scope of their national competition laws. There would seem to be a case for strengthening Article IX of GATS to deal with cartels that fall outside national jurisdictions.

3 E-commerce

There has been a particular waste of negotiating effort on e-commerce. Australia, United States and the EU, for the reasons that Patrick Messerlin mentioned, have focused on duty-free electronic commerce, which is of profound irrelevance. This is because duties are not imposed on e-commerce, and even if they were feasible, it is not at all clear that they should be prohibited when an inferior instrument, like quotas, is allowed under the GATT. Richard Snape, among others, has written of

how undesirable it is to encourage the use of quotas when the presumption should be in favour of fiscal instruments.

The focus on duties has diverted attention from some of the most important regulatory questions. Let me illustrate by one example. In some ways the possibility of e-commerce can be seen as de facto mutual recognition, insofar as it prevents countries from shutting out exports from jurisdictions that have different regulations. Because countries cannot now prevent other regulatory jurisdictions from supplying services to them, there will be greater pressure for basic harmonisation. For instance, the actions that are now being taken against financial supply sources, such as the Cayman Islands and the Bahamas, are creating pressure for basic harmonisation, because exports of financial services cannot be prevented in the way that exports of goods can be.

Reference

Fink, C., Mattoo, A. and Neagu, C. 2000, Trade in international maritime services: how much does policy matter?, Policy Research Working Paper, World Bank, Washington DC (forthcoming).

General discussion

The general discussion focused on:

- technology-driven liberalisation; and
- general versus specific approach to regulatory reform.

Technology-driven liberalisation

Richard Snape queried Patrick Messerlin's argument that liberalisation within Europe had been primarily the result of technology. The reduction of tariffs and trade barriers within the EU was not technology-driven. In relation to regulatory barriers, the key *Cassis de Dijon* decision by the European Court of Justice on mutual recognition was certainly not driven by technology. That decision had been in operation for a long time and affected services as well as goods.

Patrick Messerlin responded that his comments related to services and not to goods. While the *Cassis de Dijon* decision established the principle of cooperation and mutual recognition among Member states in the EU, its implementation had created many difficulties. The basic problem was in how to design mutual recognition agreements. The result was chaotic, and the EC had to go back to the Court of Justice in order to settle the difficulties. This had been a very long process — and was still ongoing.

General versus specific approach to regulatory reform

Richard Snape noted that David Hawes' comments illustrated how hard it was to get out of a sector-specific arrangement, such as an airline agreement, once a country was in it. The lesson was not to get into such agreements. Richard Snape agreed that the only way out of such an arrangement was through some airline-specific type of agreement. He clarified that he thought that the GATS was a remarkable achievement of the Uruguay Round. Therefore, his criticisms of it started from that point.

In reference to David Hawes comment that some airlines might go out of business because of competitive pressures, Richard Snape doubted that any significant international airline around the world had been driven out by international competition, though some had been driven out by national competitors. This illustrated the degree of insulation of the airline industry.

Christopher Findlay concluded the conference by suggesting that the conference had highlighted the fundamental dilemma about approaches to regulatory reform at

a sectoral or regional level. He agreed with Pierre Sauvé that a regional approach for like-minded countries provided opportunities to make progress on liberalisation and send messages to the broad multilateral level. However, a fundamental concern was the high incidence of regional bilateral arrangements and negotiations that were occurring in the Asia Pacific region. It may not be appropriate to pursue a regional approach to regulatory reform in that environment, which could lead to a new version of hub and spoke-ism and discrimination. It was difficult to see the regional approach as the mechanism to drive the regulatory reform process forward.

The mechanisms provided by larger groupings, such as APEC, could provide appropriate principles for regulatory reform. However, the increase in regional bilateral agreements, especially those that were based on trust rather than comparative advantage, could give a worse outcome than the current situation.

APPENDIXES

A Conference participants

Kylie Addison	Department of the Treasury
Robert Albon	Australian National University (now Australian Competition and Consumer Commission)
Kym Anderson	University of Adelaide
David Archer	Architects Accreditation Council of Australia
Steven Argy	Productivity Commission
Gary Banks	Productivity Commission
Paul Bek	Productivity Commission
Paul Belin	Productivity Commission
Graham Blinman	Productivity Commission
David Borthwick	Department of Health and Community Services
John Boshier	Institution of Engineers Australia
John Braithwaite	Australian National University
Jennifer Bryant	Productivity Commission
Alison Burrows	Department of Foreign Affairs and Trade
Neil Byron	Productivity Commission
Jeff Carmichael	Australian Prudential Regulation Authority
Richard Clarke	Productivity Commission
Paul Coghlan	Productivity Commission
John Cosgrove	Productivity Commission

Wayne Crook	Productivity Commission
Stuart Cunningham	Queensland University of Technology
John Daffey	Pharmaceutical Society of Australia
Philippa Dee	Productivity Commission
Joe Dimasi	Australian Competition and Consumer Commission
Samantha Doove	Productivity Commission
Peter Drysdale	Australian National University
Henry Ergas	Network Economics Consulting Group Pty Ltd
Christopher Findlay	Australian National University
Robert Fitzgerald AM	(NSW) Community Services Commission
Owen Gabbitas	Productivity Commission
Gerry Gentle	Productivity Commission
Ian Gibbs	Productivity Commission
Peter Greagg	Department of the Treasury
Caroline Greenway	Department of Communications, Information Technology and the Arts
David Greig	ACIL Consulting Pty Ltd
Tim Grimwade	Department of Prime Minister and Cabinet
Lisa Gropp	Productivity Commission
Kim Gusberti	Productivity Commission
Peter Harris	Department of Transport and Regional Services
David Hawes	QANTAS
Sue Holmes	Productivity Commission (now Organisation for Economic Co-operation and Development)

David Johnson	Melbourne Institute of Applied Economic and Social Research
Patrick Jomini	Productivity Commission
Stewart Jones	Department of the Treasury
David Kalisch	Department of Family and Community Services
Jeffrey Keddie	University of Melbourne
Robert Kerr	Productivity Commission
Roger Kerr	New Zealand Business Roundtable
Stephen King	University of Melbourne
Helen Lapsley	University of New South Wales
Ralph Lattimore	Productivity Commission
Tamara Lee	Productivity Commission
Greg McGuire	Productivity Commission
Will Martin	World Bank
Aaditya Mattoo	World Bank
Patrick Messerlin	Institut d'Etudes Politiques de Paris
Jonas Mockonas	Department of Prime Minister and Cabinet
Alan Moran	Institute of Public Affairs
Duc Nguyen-Hong	Productivity Commission
Jane O'Donohue	Productivity Commission
Helen Owens	Productivity Commission
Paul Palisi	Department of Prime Minister and Cabinet
Franco Papandrea	University of Canberra
Brian Perkins	AAPT

Mark Pierce	Department of Foreign Affairs and Trade
Garth Pitkethly	Productivity Commission
Gerard Prior	Productivity Commission
Radmila Ristic	Productivity Commission
David Rivett	Australian Medical Association
Stephen Rowley	Productivity Commission
Graeme Samuel	National Competition Council
Pierre Sauvé	Harvard University (now Organisation for Economic Co-operation and Development)
Robyn Sheen	Productivity Commission
Deena Shiff	Telstra
Rod Shogren	Australian Competition and Consumer Commission
Tony Slatyer	Bureau of Transport Economics
Judith Sloan	Productivity Commission
Richard Snape	Productivity Commission
Duncan Spender	Department of the Treasury
Andrew Stoeckel	Centre for International Economics
Ruth Thomson	Productivity Commission
Anne Trimmer	Law Council of Australia
Tom Valentine	University of Western Sydney
George Verikios	Productivity Commission
Tony Warren	Network Economics Consulting Group Pty Ltd
Ed Willett	National Competition Council

John Williams	Productivity Commission
Lynne Williams	Productivity Commission
Ross Wilson	Productivity Commission
Mike Woods	Productivity Commission

B Conference program

Day 1 — Monday 26 June 2000

8.30 – 8.45 Registration

8.45 – 9.00 **Introduction**
Gary Banks (Productivity Commission)

9.00 – 10.30 Session 1: Why regulate services?

Introduction to the regulation of services
Christopher Findlay (Australian National University)

Discussant:
Patrick Messerlin (Institut d'Etudes Politiques de Paris)

The principles of good regulation
Paul Coghlan (Productivity Commission)

Discussant:
Aaditya Mattoo (World Bank)

10.30 – 11.00 Morning Tea

11.00 – 12.15 Session 2: Regulating for competition — rail,
telecommunications and air transport

Access: what, where and how
Stephen King (University of Melbourne)

Discussant:
Ed Willett (National Competition Council)

Regulating access to telecommunications networks

Robert Albon (Australian National University)

Discussant:

Rod Shogren (Australian Competition and Consumer Commission)

12.15 – 1.45 Lunch

1.45 – 2.30 Session 2: (continued)

Improving the regulation of international air services

Geraldine Gentle (Productivity Commission)

Discussant:

Christopher Findlay (Australian National University)

2.30 – 3.45 Session 3: The great debate

Is market-based regulation is the best way to regulate community services?

Debaters:

Alan Moran (Institute of Public Affairs)

Robert Fitzgerald AM (Community Services Commission)

Mediator:

Helen Owens (Productivity Commission)

3.45 – 4.15 Afternoon Tea

4.15 – 5.30 Session 4: Regulating for consumer confidence — finance, the professions and pharmacies

Lessons from the Asian Crisis

Jeff Carmichael (Australian Prudential Regulation Authority)

Discussant:

Tom Valentine (University of Western Sydney)

Principles for regulating the professions

Neil Byron (Productivity Commission)

Discussant:

Jeffrey Keddie (University of Melbourne)

Day 2 — Tuesday 27 June 2000

9.00 – 9.45 Session 4: (continued)

Regulating for the safe and effective delivery of pharmacy services

David Greig (ACIL Consulting Pty Ltd)

Discussant:

Helen Lapsley (University of New South Wales)

9.45 – 10.30 Session 5: Regulating to attain social goals

Regulating gambling: a ‘market friendly’ approach to the social impacts

Gary Banks (Productivity Commission)

Discussant:

David Johnson (Melbourne Institute of Applied Economic and Social Research)

10.30 – 11.00 Morning Tea

11.00 – 12.15 Session 5: (continued)

Regulating culture: has it ‘Gone with the Wind’?

Patrick Messerlin (Institut d’Etudes Politiques de Paris)

Discussant:

Pierre Sauvé (Harvard University)

Towards multilateral rules on trade and culture: protective regulation or efficient protection?

Pierre Sauvé (Harvard University)

Discussant:

Patrick Messerlin (Institut d’Etudes Politiques de Paris)

12.15 – 1.45 Lunch

1.45 – 3.00 Session 6: What are the costs of too much regulation?:
measuring and modelling the impact of restrictions on
trade in services

**‘Sins of commission and omission’: measuring regulatory
impediments to trade in services**

Tony Warren (Australian National University)

Discussant:

Alison Burrows (Department of Foreign Affairs and Trade)

Modelling the liberalisation of services

Philippa Dee (Productivity Commission)

Discussants:

Kym Anderson (University of Adelaide)

Will Martin (World Bank)

3.00 – 3.30 Afternoon Tea

3.30 – 5.00 Session 7: Achieving better regulation of services

**Approaches to improving regulation: regulatory competition or
regulatory harmonisation? A forum**

Chair and Panellist:

Christopher Findlay (Australian National University)

Panellists:

Kym Anderson (University of Adelaide)

David Hawes (Qantas)

Will Martin (World Bank)

Aaditya Mattoo (World Bank)

Patrick Messerlin (Institut d’Etudes Politiques de Paris)

Alan Moran (Institute of Public Affairs)

Pierre Sauvé (Harvard University)

Richard Snape (Productivity Commission)