The Poverty Line Revisited

James Cox

Poverty has been in the news recently. According to a report released in December 2001 by the Smith Family and the National Centre for Social and Economic Modelling (Natsem), ‘in 2000, about one in every eight Australians lived in income poverty — 13 per cent of all Australians’ (p. vii). Moreover, ‘there was a steady increase in poverty among all Australians during the 1990s’ (Harding, Lloyd and Greenwell, 2001:4).

A response was provided by a team from the Centre for Independent Studies (CIS) in January 2002 (Tsumori, Saunders and Hughes, 2002). The CIS authors concluded that the findings of the Smith Family/Natsem report are dubious because they confuse poverty and inequality and ‘rest on data that is in some respects unreliable’(p. 1). They argue that ‘the debate over the future of welfare requires more reliable data on the extent and causes of deprivation’. Other commentators have subsequently joined in the discussion.

Controversy over poverty lines is nothing new in Australia. The Smith Family/Natsem report uses a modified form of the Henderson poverty line. This poverty line was developed for a Melbourne study in the late 1960s and rose to national prominence in the report of the Commission of Inquiry into Poverty in 1975. This poverty line has been subsequently updated by the Melbourne Institute according to changes in Average Weekly Earnings and (later) Household Disposable Income per capita. The Henderson poverty line was never endorsed by government and, from time to time, criticisms of the Henderson poverty line were mounted from official sources. Eventually, the Minister for Social Security asked the Social Welfare Policy Secretariat to construct a poverty line that would be suitable for use in contemporary Australia.

The Social Welfare Policy Secretariat (SWPS, 1981) subsequently reported that the task was impossible. They concluded (p. 208) that:

there are no entirely satisfactory techniques which would allow us to assess whether the money amounts of pensions and benefits are adequate or not. Further, many factors other than the needs of families in different circumstances influence decisions about the level of pensions and benefits.

I was the principal author of the SWPS report. Despite, or perhaps because of this finding, the Henderson poverty line has continued to be published. Although interest may have diminished in the years following publication of the SWPS report, more recently the poverty line has made a comeback. Despite the

James Cox is a member of the New South Wales Pricing Tribunal.
absence of official endorsement, the Henderson poverty line has, for example, been quoted in ABS publications (see, for example, ABS, 1998:125-129).

In asking me to write this article, the editors of Agenda have kindly invited me to revisit these issues twenty years later. My purpose in writing this article is to investigate the usefulness of the estimates of the extent of poverty that have recently gained such prominence in the Australian debate and to comment on the work that needs to be done if we are to understand better the extent and nature of disadvantage in Australia today. I will proceed by first reviewing the recent contributions by the Smith Family/Natsem authors and the CIS authors. I will then discuss the issues that are raised by these authors. These include: what is meant by poverty in Australia today; the problems in representing these ideas about poverty by a poverty line; and the usefulness for policy of the estimates by the Smith Family and Natsem of the composition of the low income population. Finally, some conclusions are drawn.

The Smith Family/Natsem Report

This report was commissioned by the Smith Family and prepared by three authors who work at the National Centre for Social and Economic Modelling at the University of Canberra. In her forward, the Director of Natsem refers to ‘The Smith Family’s report on financial disadvantage and poverty in Australia’. Certain critical judgements (such as the choice of a poverty line based on half average income) are justified, according to the authors, by ‘The Smith Family’s commitment to work for a more cohesive society’. However, the Chief Executive Officer of the Smith Family refers to the ‘annual Smith Family/Natsem report on financial disadvantage’. It seems safest to conclude that both organisations are jointly responsible for the whole report.

The authors of the Smith Family/Natsem report develop their arguments through a number of stages. First, they adopt a relative definition of poverty. This is controversial, and is discussed further below. Secondly, the authors develop a number of alternative poverty lines. Twelve poverty lines are constructed. These differ accordingly to the method of setting the line (half average income, half median income or the ‘traditional Henderson’ approach), the equivalence scale used to adjust for differences in families’ needs, the method of updating the poverty line and whether poverty before or after housing costs is considered. As might be expected, these poverty lines differ greatly. The authors point out (p. 24) that all but one of these poverty lines suggest that poverty rose during the 1990s. However, poverty lines based on half median income show no increase in poverty, or only a small increase, over the 1990s.

The third step taken by the Smith Family/Natsem authors is to discuss in some detail the methodological choices involved in developing a poverty line. The problems and choices made by the authors are set out carefully. Indeed, it could be argued that the Smith Family/Natsem report includes a criticism of itself that anticipates many of the points that are made by the CIS authors. Having discussed the problems in setting poverty lines the Smith Family/Natsem authors
conclude that ‘it is probably safer to discuss “financial disadvantage” rather than “poverty”’ (p. 27). Sometimes these authors refer to ‘income poverty’. This suggests that there are other aspects of poverty that do not depend on income and cannot be represented by the poverty line.

The fourth step taken by the Natsem/Smith Family authors is to choose a preferred poverty line. This line is set for a standard family of two adults and two children at half average disposable income (that is, income after income tax has been paid). The incomes of other families are made comparable with those of standard families through adjustment for differences in need by the Henderson equivalence scale. This fourth step is controversial. As the CIS authors correctly note, most researchers these days use relative poverty lines that are based on median rather than average income. Median incomes are less affected than average incomes by changes in very high or very low incomes that are not representative of the incomes earned by most people. Average incomes grew more rapidly over the 1990s than median incomes because high incomes grew rapidly during the decade.

The Smith Family/Natsem authors do not provide an extensive justification for their choice of poverty line. They argue (p. 2) that a poverty line defined in terms of average income is transparent and easy to update through time. There is, they argue, no clear consensus about whether it is preferable to take a proportion of median or average income as the poverty line (p. 26). Moreover, use of the half average income line ‘denotes acceptance of the proposition that the living standards of the poor should be measured against the living standards of all the population, including the affluent’. Use of a poverty line based on half average income ‘better captures relative deprivation in times of rising income inequality’ and therefore reflects ‘the Smith Family’s commitment to work for a more cohesive society’ (p. 27).

These arguments are not convincing. Although it may be of interest to compare the living standards of the poor with those of the affluent, this is not the same thing as defining poverty partly in terms of the living standards of the affluent. If poverty is not being able to afford to purchase the goods and services that most people enjoy, it is the living standards of ordinary people that is relevant. Most people cannot afford to buy the luxuries that the affluent enjoy. Nor is it necessarily true that a society where income is equally distributed will be cohesive. A country may be able to achieve a fairly equal distribution of income through regulation of the labour market and provision of generous welfare payments for people who are unable to find employment at the regulated wages. (Some European countries are like this.) Such a country is likely to be sharply divided between those who are working and those who are receiving benefits. Those who are not able to find work at the regulated wage are likely to resent their exclusion from work; those who are working may well resent the amount they have to pay in tax to support people who may not be much worse off than they are. By contrast, a less egalitarian society with higher levels of employment may well be more cohesive.
Having chosen a poverty line, the fifth step undertaken by the Smith Family/Natsem authors is extensive analysis of the characteristics of groups with incomes less than the poverty line and how this has changed through time. This analysis is of considerable interest even for those who do not accept the Smith Family/Natsem poverty line; it provides information on the lowest 10 to 15 per cent of the population by income. I discuss this analysis later in this article.

Finally, the Smith Family/Natsem authors emphasise the headline conclusions about poverty and its growth that were outlined at the start of this article. They note (p. 24) that:

- given the strong link between unemployment and poverty it is of concern that our society did not manage to reduce poverty during a period of strong economic growth. As we enter more uncertain economic times, the need to protect those at the bottom may become greater.

The CIS Report

A team of authors from the Centre for Independent Studies has provided a commentary on the Smith Family/Natsem report. They conclude (p. 2) that the Smith Family/Natsem report

- has been professionally researched and painstakingly compiled and it has received a lot of attention in the media. Its conclusions are, however, misleading.

The CIS authors reach this view for three main reasons. First, in the opinion of the CIS authors the Smith Family and Natsem have confused poverty and inequality by adopting a relative view of poverty. By contrast, most people think of poverty in absolute terms as a shortage of food, accommodation and clothing. The CIS authors (p. 1) argue that:

- sustained economic growth coupled with government welfare provision means that absolute poverty is extremely rare in Australia today. Indeed, as the Smith Family report itself notes, Australia’s poorest families became better off in absolute terms during the 1990s — their real (inflation-adjusted) incomes rose by an average of $38 per week over this period.

Second, the CIS authors (p. 3) are of the view that, by adopting a poverty line based on average rather than median income, the Smith Family and Natsem have made an unusual choice that exaggerates the seriousness of poverty in Australia.

- Had the more usual median income been used as the basis for these calculations, the result would have been (a) a much lower estimate of the size of the poverty problem (8.7 per cent as compared with 13 per cent),
and (b) a conclusion that poverty did not worsen over the Howard years. This would have given us a different (and much less dramatic) picture of the scale and nature of ‘relative poverty’ in Australia.

Third, the CIS authors argue that the Australian Bureau of Statistics (ABS) data on which the Smith Family and Natsem rely are not sufficiently reliable for conclusions to be drawn from them about the extent of poverty. The problems include the under-reporting of incomes especially by welfare beneficiaries and the self-employed, the exclusion of certain important sources of income (such as income in kind and government benefits provided below cost) and the decision by the Smith Family/Natsem authors to base their estimates on weekly income. Some families that had an unusually low income in the survey week may not have been counted as poor had data been collected for a longer period such as a year. There are, moreover, many problems involved in choosing an equivalence scale to take account of the differing needs of households of different size and composition. Thus (p. 10):

the income data that get analysed in studies such as this one are constructions of the researchers themselves. It is quite possible to shift estimates of poverty up or down simply by tweaking the assumptions you make about the equivalent running costs of different types of household.

Finally, the CIS authors argue that expenditure data provide a better basis for assessing standards of living than income data.

Despite their scepticism, the CIS authors (p. 11) conclude by arguing that:

the development of reliable poverty estimates is an essential precondition for any reasoned debate about the direction that social policy in Australia should take over the years to come.

They think that this can be done by developing ‘an absolute poverty line adjusted for different kinds of households and location’ based on calculating the costs of necessities. They acknowledge that this is a difficult exercise.

I think that it is useful to comment on a number of aspects of this debate:

- What is meant by poverty and how well the concept of poverty is represented by absolute or relative poverty lines;
- The methodological problems in constructing poverty lines;
- The continuing attractiveness of poverty lines, despite all the problems;
- The conclusions of the Smith Family/Natsem report on the composition of the low income population and the implications of these findings, if any, for government policy;
• What can most usefully be done to further advance an understanding of poverty and deprivation in Australia.

Comments follow on each of these points.

The Concept of Poverty

Poverty is usually taken to mean the absence of the essentials of life, such as food, shelter and clothing. It is a disgraceful state: poor people lack, in Adam Smith’s words the ability to appear in public without shame. Frequently it is argued that it is the responsibility of government to ensure that people can avoid the disgraceful state of poverty.

I think it is reasonable to argue that the essentials of life include more than just certain goods and services. This focus on commodities ignores, at a time when despair and hopelessness seem to be so common, less tangible aspects of living decently. Children do not always have a safe and stable environment in which they can develop easily the skills required for success in adult life, for example, because of widespread substance abuse within their communities. Adults do not always live in an environment in which it makes sense for them to believe that they can improve their circumstances through their own efforts, for example, because of widespread welfare dependency. The absence of hope only too often leads to self-destructive behaviour, violence and suicide. Noel Pearson and others have recently reminded us that these are serious problems in some remote aboriginal communities (Pearson, 2001). Although these communities face particular difficulties, these problems are not confined to them.

The CIS authors argue that absolute poverty (in the sense of having insufficient income to maintain a minimum standard of life) is extremely rare in Australia today. This statement is correct as far as it goes. However, because of the behavioural problems noted above, not everyone in Australia has everything they need to live a decent life. These non-material aspects need to be considered, it seems to me, in a full account of absolute poverty in Australia.

Some recent writers have attempted to develop an understanding of poverty that is more profound than mere absence of commodities. For example, Amartya Sen has encouraged us to think about a person’s well being in terms of his or her ‘capability set’ — the opportunities potentially open to the person as well as the one actually chosen (see for example, Sen, 1999:75). Poverty may then be thought of as capability deprivation. The capability deprivation approach is readily able to encompass non-material aspects of poverty. Although low income is a source of capability deprivation, so too may be lack of access to education and health services and unemployment rates that ‘make a mockery of people’s ability to help themselves’ (Sen 1999:98). Moreover, the relationship between income and capability deprivation is variable, depending on the circumstances of communities and families within those communities. For example, a higher income is needed to sustain a decent standard of living in a richer community than in a poorer one because what is considered decent varies with the standard of
living of the community. This provides some support for the idea of relative poverty.

However, the association between poverty and the poverty line breaks down if this more sophisticated understanding of poverty is adopted. The circumstances of families vary greatly. As the early poverty researchers noted, families vary greatly in their ability to manage on a limited income. Some families are able to grow their own food or receive other types of income in-kind. Some families own houses or other assets that reduce the amount they would otherwise have to spend, for example, on rent.

A family in which someone cooks, cleans and attends to household repairs will enjoy a higher standard of living than a household in which no-one does these things. Travers and Richardson (1993:201-202) writing about a study of material well being in Australia that they had undertaken note that:

in particular, we found that inequality in money income and inequality in other components of full income (for example, home-ownership) tend to offset each other. The use of full income does not produce a more egalitarian picture of Australia than does current income, but it does show that there is little compounding of disadvantage when the broader aspects of material well being are considered. Those who are weak in one resource (income) are often strong in another (housing or non-employed time).

Moreover, disadvantage in material well being tended only to a limited degree to spill over into other aspects of life such as social activities, happiness, health and a general sense of optimism. According to Travers and Richardson (p. 206):

it requires a great deal of information about a person’s assets, family relationships, receipts in kind and income before his or her material well being can be identified with any accuracy.

All this calls into question whether what we mean by poverty can be adequately represented by any poverty line.

**Problems in Constructing Poverty Lines**

In my view the issues outlined above are sufficiently important to make it unlikely that a poverty line can be developed that successfully separates those people who are able to live decently from those who are not able to do so. However, this is not the view of either the CIS authors or of the Smith Family/Natsem authors.

The CIS authors argue that it is possible to develop an absolute poverty line for Australia (p. 11). This involves first developing a list of goods and services for different types of families that are essential in Australia today and then calculating an amount of income that is sufficient to purchase these essentials. The poverty
lines that were developed by the early poverty researchers in late nineteenth and early twentieth century England emphasised what was required for ‘mere physical efficiency’.

I think it is unlikely that anyone would advocate so austere a poverty line in Australia today. Instead a researcher would have to consider questions such as whether not eating meat or not having a telephone, car, television, holiday or an occasional night out, constitutes poverty. The difficulties are considerable. Not everyone wants to eat meat, view television or drive a car. Thus it is hard to know (in the absence of a good deal of information about a family and the community they live in) whether the absence of these things indicates voluntary choice by a household or deprivation. The early poverty researchers (such as Booth and Rowntree in England) were able to draw on a great deal of detailed knowledge about the communities they were studying in developing the first poverty lines and making estimates of the number of people in those communities with incomes below these lines. By contrast, who from the relative (and remote) comfort of a research institution or government department would be so presumptuous as to specify the minimum requirements of a decent life in contemporary Australia?

The CIS authors draw attention to a low cost budget standard constructed by researchers at the Social Policy Research Centre (SPRC). As noted in a SPRC discussion paper (Saunders, 1998:6), this low cost budget standard represents a standard of living which may require frugal and careful management of resources but would still allow social and economic participation consistent with community standards and enable the individual to fulfil community expectations in the workplace, at home and in the community.

The SPRC researchers emphasise that the low cost budget standard is not a poverty line; it is closer to 70 to 75 per cent of median living standards than to 50 per cent.

The SPRC discussion paper (Saunders, 1998:15) noted the difficulty of reducing this budget standard to the poverty level:

(To) reduce the low cost standard for a couple with two children to one half of the corresponding modest but adequate standard would involve reducing it from $602.10 (a week) to $408.70, or by $193.40. For this family type, a budget of $408.70 would not allow them to meet their housing, energy, food and clothing needs at the low cost standard .... There would thus be no money left over for health, transport, leisure and personal care needs — all of which facilitate social participation — nor would there be any money for household goods and services, a category which includes expenditure on many basic household goods as well as education and child care costs.
All participants in this debate accept that researchers have a good deal of discretion in developing relative poverty lines. Although the choices made by the Smith Family/Natsem authors are reasonable, other equally reasonable choices could have been made that would have resulted in very different answers. The main issues here include the following:

- ABS income surveys are the main source of data for the Smith Family/Natsem study. The institutional and homeless populations, which are perhaps particularly likely to be disadvantaged, are excluded from the survey population. The income survey also suffers from sampling error. The Smith Family/Natsem authors think that the 1997 survey is so unreliable (it shows a lower level of poverty than the preceding and subsequent surveys) that they have excluded it from the analysis (p. 51). However, this is a somewhat arbitrary approach to the problem.

- Some sources of income are under-reported in the surveys (for example, income from self employment). Moreover, income as reported in the surveys grew less rapidly than the broader measures of income that are presented in the National Accounts.

- The percentage of the population that is estimated to be in poverty depends on whether weekly or annual income is measured, and on whether the household or the nuclear family is chosen as the unit of analysis. The Smith Family/Natsem authors have chosen procedures that lead to larger estimates of the extent of poverty than some alternative procedures. Because weekly income is more variable than annual income, poverty appears to be greater if weekly income is used. Because nuclear families (or abbreviations of them such as adult children) may share resources within a household, poverty appears to be greater if the family rather than the household is used as the unit of analysis.

- Some decisions made by the Smith Family/Natsem authors or the ABS about particular families can be questioned. For example, the income of families with negative incomes were reset to zero. Income tax was imputed by the ABS on the basis of income and family structure rather than being reported directly.

- As noted earlier, the basis for setting a relative poverty line is entirely arbitrary. Opinions differ as to whether median income, average income or some other basis (for example, a minimum wage) should be used to construct a poverty line. Choosing 50 per cent of (say) median income as the poverty line is arbitrary. Moreover, the percentage of the population that appears to be in poverty is very different if a poverty line is set at 40 or 60 per cent of median incomes, not 50 per cent. However, there is no clear basis for making these choices.

- Relative poverty lines incorporate equivalence scales which show how much more or less income a non-standard family needs to achieve the same standard of living as a standard family.
Developing equivalence scales is a difficult task both in principle and in practice. The equivalence scales that are used by the Smith Family/Natsem authors have been chosen because they appear to be reasonable but they have no clear empirical basis. By contrast, equivalence scales that do have some empirical basis (either based on household expenditure survey data or surveys that request people to make subjective assessments about their own living standards) typically show too small an addition for additional household members to be intuitively appealing — see Buhmann et al. (1988) for further discussion of this issue. However, particularly in a country such as Australia where average household size is fairly large, estimates of the extent of poverty depend greatly on the choice of the equivalence scale.

The Continuing Popularity of Relative Poverty Lines

The points made above are not new and, indeed, are accepted both by the Smith Family/Natsem authors and by the CIS authors. Why, then, do researchers continue to make so much use of relative poverty lines? I think two points are worth making here.

First, relative poverty lines are easy to develop. Governments collect a good deal of information on incomes that is made available to researchers. There is no need to develop a detailed understanding of the circumstances of local communities, as the early poverty researchers were required to do. Moreover, there is no need to make explicit (and difficult) judgements about which goods and services are required for a decent life.

A second reason is the effectiveness of claims about the number of people in poverty in arguing for increases in government benefits. Under the Howard government, unemployment benefits have increased less rapidly than other social security payments (see the Smith Family/Natsem report, p. 23). It is not unreasonable to argue that unemployment benefits should be increased (although the effect of any increases on work incentives should also be considered). However, the argument that benefits should be increased to reduce the number of people in poverty has greater rhetorical force than arguments about the weight that should be given in making policy decisions to each of the competing, but desirable, objectives of improving the adequacy of benefits and minimising work disincentives.

The Low Income Population

The Smith Family/Natsem report includes an interesting discussion of changes in the composition of the low income population during the 1990s. The authors refer to this group as being in poverty or income poverty. However, even people who do not accept the Smith Family/Natsem poverty line may find the analysis of some interest because it describes the characteristics of families who receive the lowest incomes. Because the Smith Family/Natsem authors use a poverty line that increases in line with average income, the percentage of families with incomes
below the line grew over the 1990s. Although this is unlikely to have been a major factor, it does need to be considered when reading the results of the analysis.

The main findings of the Smith Family/Natsem authors regarding the composition of the low income population include the following:

- Sole parents and single people are more likely than average to be members of the low income population; couples with or without children are less likely than average to be members of the low income population.
- Persons whose principal source of income is government benefits are more likely than average to be members of the low income population; persons whose main source of income is wages or salary are less likely than average to be members of the low income population.
- Families in which there are no earners or one part-time earner are more likely than average to be members of the low income population; families in which there is at least one full-time earner are less likely than average to be members of the low-income population.
- Persons without post-school qualifications are more likely than average to be members of the low income population; persons with such qualifications are less likely than average to be members of the low income population.

The main findings regarding changes in the composition of the low income group include the following:

- Sole parents were less likely to be members of the low income group at the end of the 1990s than at the beginning of the decade. Single people were more likely to be members of the low income group at the end of the decade than at its beginning.
- The percentage of children who are in the low income group fell during the early part of the 1990s but then increased.
- The percentage of aged who are in the low income group grew during the 1990s (but remained below the average for the population as a whole).
- Persons receiving government benefits and part-time earners made up a larger proportion of the low income population at the end of the 1990s than at the beginning of the decade.

Based on these findings, the best advice you could give to someone to prevent them from becoming a member of the low income population is to receive an education, work in a full time job and to marry and to remain married. In the preface to the Smith Family/Natsem report, two Smith Family authors rightly emphasise the importance of education but do not discuss the importance of marriage or full time work.

During the 1990s spending on income support by Australian governments increased from 7.3 per cent of GDP in 1989/90 to 9.8 per cent of GDP in 1999/2000 (ABS, 1996; ABS 2001a). The percentage of the population receiving
income support (and especially the percentage of the population of working age) also increased over the period (see ABS, 2001b:166-169) although economic and demographic conditions were generally favourable. Despite this increase in spending the percentage of the population with incomes below the Smith Family/Natsem poverty line was either stable or increasing.

There is much evidence to suggest that the number of people receiving income support payments depends, among other things, on the level and generosity of those benefits. It is worrying therefore, that the structure of incentives provided by Australia’s social security programs encourages sole parenthood over the two-parent family and encourages the receipt of benefits (or a combination of benefits and part time work) over full-time work. Although income support programs are important in avoiding hardship, it is also evident that, at the very least, they are not easy answers to the problems that they seek to address. Moreover, there is a voluntary aspect to poverty if people respond to the incentives provided by the social security system. In terms of the language of the capabilities approach, the opportunities (or capability set) available to those currently below the poverty line may extend above it (although not always by very far). This also needs to be considered when thinking about poverty in Australia.

**Concluding Thoughts**

The argument of this article has been that poverty lines are far too simple and arbitrary to represent clearly the disadvantage that undoubtedly exists in Australia today. We need to consider more than just income and understand more about the problems that particular communities experience. The ease of construction and rhetorical effectiveness of relative poverty lines will ensure that they will continue to be used. I hope, however, that the arguments made in studies such as this one will encourage the development of a better understanding of what is meant by poverty and disadvantage and further studies that consider more than a short-term measure of income in assessing the standard of living of families and households.

**References**


The author gratefully acknowledges helpful comments by two anonymous referees.
Copyright Law and Competition Policy: International Aspects

Theo Papadopoulos

A recent Federal Court Ruling in a case brought by the Australian Competition and Consumer Commission (ACCC) against a number of record companies, alleging breaches of the Trade Practices Act 1974 (TPA), highlights a growing interconnection between international trade in copyright product, copyright law and competition policy. The breaches were symptomatic of unwillingness on the part of foreign copyright owners (acting through their respective territorial licensees) to accept intra-brand competition brought about by the Copyright Amendment Act (No 2) 1998 that permitted the parallel importation of non-infringing copies of sound recordings. Strategies to retain copyright owner control over vertical distribution included attempts to block both parallel exports and parallel imports. The interplay between the economics and law of trade in copyright product is a complex area that has until recently attracted little attention by economists (Fels, 2001).

A range of issues highlighted in the above Federal Court Ruling is explored in this paper. The issues addressed in the paper include the international dimensions of intellectual property law, vertical restraints in the international distribution of copyright product and the implications for domestic competition policy.

Authorised and Unauthorised Distribution Channels

The Copyright Act 1968 bestows a number of exclusive rights on creators of literary, musical and artistic works, including the right to reproduce, to perform, to adapt and to translate the work. These economic rights are necessary to promote creative endeavour and to avoid imitation (free riding) and market failure. In determining the range and strength of these monopoly rights, policy makers need be mindful of the potentially anti-competitive consequences that may ensue.

Copyright is divisible by time, act and territory. Exclusive territorial licenses, contractual arrangements negotiated between copyright owners and domestic agents within specific national territories, are a form of vertical restraint that enable copyright owners to coordinate the international distribution of copyright product. The result is a series of territory based distribution channels within which a particular copyright product is exclusively distributed by the copyright owner or his/her licensee.

Theo Papadopoulos is Senior Lecturer in the School of Applied Economics, Victoria University (Melbourne) where he coordinates the Bachelor of Business — Music Industry degree program.
In exchange for a license fee, the territorial licensee obtains exclusive rights to the distribution of the copyright product within the territory. This provides the necessary security for licensee pre-sales (marketing and promotion) and post-sales (servicing and warranty) investments. Distributors of competing copyright product likewise typically enjoy exclusive territorial rights. To the extent that this encourages investment and product variety, exclusive territorial licenses promote inter-brand rivalry and are pro-competitive. On the other hand, this exclusivity produces local monopolies for specific copyright product and is not unambiguously welfare enhancing.

Parallel imports refer to the importation of a copyright product (released in a foreign territory) by someone other than the domestic license holder. Parallel imports arise where enterprising traders exploit territorial price differentials, diverting copyright product from a low-price territory to a high-price territory (Moran, 1999). Territorial price differentials arise for a host of reasons including exchange rate movements, international price discrimination, national price regulations, vertical controls and variations in pre-sales and post-sales investments (Maskus and Lahouel, 1999).

Parallel imports provide an unauthorised distribution channel that introduces intra-brand competition, and in the view of copyright owners and licensees, is a disruptive element to an otherwise efficient distribution system. Motivations cited for copyright owner control over parallel imports include the prevention of free riding, maintenance of product quality, detection of illicit copies (piracy), price discrimination and collusion (Chard, 1989; Abbott, 1998; Gallini, 1999).

While collusion is anti-competitive and likely to lower social efficiency, the welfare consequences of international price discrimination are less clear. Under certain assumptions, international price discrimination, backed by parallel import restrictions, can increase global welfare (Malueg and Schwartz, 1994). However, for a net importer of copyright product like Australia, the introduction of competition from parallel imports can enhance national welfare. Indeed, depending on the response by copyright owners in low price territories from which parallel imports are diverted, competition from parallel imports may even increase global welfare (Papadopoulos, 2000).

The remaining motivations are pro-competitive and likely to improve private efficiency. The debate over the legality of parallel trade centres on the issue of the timing of copyright exhaustion.

Exhaustion of Copyright and International Intellectual Property Law

The exhaustion of copyright lies at the heart of a contentious and unresolved issue in international intellectual property law. Copyright in most countries provides creators with a bundle of exclusive rights, including the right to make the product available on the market for the first time. This distribution right is referred to as the doctrine of first-sale. In the context of international trade in copyright product, the distribution right has created great uncertainty for copyright owners, licensees, importers, distributors and retailers alike. The controversy surrounds the issue of
whether the copyright owner should control distribution of a particular copy of a product beyond the first sale. The principle of international exhaustion holds that the distribution right is exhausted once a particular copy of the product is sold anywhere in the world. In other words, the right to distribute a specific copy of the copyright product (for example, a sound recording title, book title or business software program) is exhausted after the first sale and the copyright owner can no longer control distribution. The copyright owner cannot thereafter prevent the importation of that product into another territory.

The adoption of the principle of international exhaustion would allow the parallel importation of legitimate copies of a copyright product released onto the market anywhere in the world. The principle of national exhaustion holds that the distribution right is exhausted only within the country in which the product is first sold. The distribution right survives until the sale of that same copy in another country. This amounts to an importation right and a prohibition on parallel imports. Copyright product can only be legally imported by the copyright owners or licensees or with their expressed authorisation.

An importation right, therefore, provides copyright owners with a statutory monopoly over the importation and distribution of copyright product. In this way, the right is a form of vertical restraint that enables copyright owners to control the international distribution of copyright product. It has been argued that this statutory monopoly provides copyright owners the wherewithal to partition the global market into national segments and implement a strategy of price discrimination reflecting varying incomes and elasticity of demand (Malueg and Schwartz, 1994; Papadopoulos, 2000).

An importation right provides copyright owners with greater market power and control and consequently, enhances the economic value of the copyright. Not surprisingly then, national governments in countries that are net exporters of copyright product advocate stronger economic rights, including a prohibition on parallel imports. Intellectual property owners in these countries have urged government to advocate the adoption of the principle of national exhaustion in forums such as the WTO.

Distribution is a key right identified in the World Copyright Treaty (WCT) and World Phonograms and Performers Treaty (WPPT) and bestows on producers the exclusive right of making copies of their product available to the public. Interestingly, article 6 of the WCT explicitly avoids determination with respect to the timing of the exhaustion of this right.

Nothing in this Treaty shall effect the freedom of the Contracting Parties to determine the conditions, if any, under which the exhaustion of the right in paragraph (1) applies after the first sale or other transfer of ownership of the original or a copy of the literary and artistic works with the authorisation of the producer of the literary and artistic works.
In other words, members of these copyright conventions are free to determine the timing of the exhaustion of the right of distribution and adopt the principle of national or international exhaustion as they see fit.

Negotiations within the World Trade Organisation (WTO), as evidenced in the text of the Trade Related Intellectual Property Rights Agreement (TRIPS), have also left the controversial issue of the importation right up to individual national regulators. Article 6 of TRIPS (WTO, 1995) deals with the issue of exhaustion in the following way:

For the purposes of dispute settlement under this Agreement, subject to the provisions of Articles 3 and 4, nothing in this Agreement shall be used to address the issue of the exhaustion of intellectual property rights.

There was considerable disagreement on this issue at the time of its drafting. Ultimately, trade negotiators agreed to disagree and this question was left to individual member states to determine their respective policy positions. In short, there is nothing in international trade and intellectual property rights (IPR) law that mandates the provision of an importation right. As a statutory impediment to the free flow of copyright product, an importation right would seem to conflict with the underlying objective of the WTO: to promote free trade via the removal of tariff and non-tariff barriers to trade.

Trade Related Copyright Law Reform in Australia

Australian copyright law reform, and the liberalisation of trade in copyright product, has progressed in a step-wise fashion for individual product classes. A review of the book market in 1992, by the then Prices Surveillance Authority (PSA), resulted in a partial removal of the importation right. Parallel importation is now permissible, for example, if a book is not published in Australia within 30 days of its first publication overseas.

A review of the market for sound recordings (PSA, 1990; Senate Legal and Constitutional Legislation Committee, 1998) culminated in the Copyright Amendment Bill (No.2) 1998, permitting the parallel importation of non-infringing copies of a sound recording. More recently, a review of all remaining copyright related products saw the drafting of the Copyright Amendment Bill (2001).

The Copyright Amendment Bill proposes to remove restrictions on the parallel importation of all copyright products, including books but exempting motion pictures. The House of Representatives passed the Bill in June 2000 but it lapsed with the November 2001 election. The (slightly amended) Bill was reintroduced in March 2002. Its passage through the Senate is uncertain since both the Labor opposition and Democrats have expressed their opposition to the

1 A ‘non-infringing’ copy is defined as one that does not breach copyright law in the country of manufacture and is released into that market with the owner’s consent.
Bill. For the moment, at least, it remains illegal to import business software, video games, periodicals and similar products without the permission of the copyright owner or their domestic licensee.

The Australian government has adopted a rule of reason approach to copyright law reform and parallel imports. It has undertaken a sequential review and economic cost-benefit analysis of each category of copyright product and amending copyright law accordingly.

For example, the proposed exemption of motion pictures relates to the unique characteristics of that product and the prevailing market dynamics. Release timing is a critical dimension of competition and distribution in this market. The primary income earning activity is the public performance. A new motion picture is first released in movie theatres, and then to video rental stores before ultimately being released on DVD and videocassette. Release timing varies between territories and is a function of, among other things, seasonal factors. The removal of parallel imports could see the importation of DVD and videocassette copies of films prior to their scheduled release in the domestic market with obvious financial costs to both copyright owners and the movie theatre industry. It makes sense, therefore, to adopt the principle of national exhaustion with respect to this product class.

Not surprisingly, given the economic significance of copyright dependent industries and the income redistributive effects of changes to the nature and extent of exclusive commercial rights, copyright law reform has been very controversial. The response to the adoption of international exhaustion with respect to sound recordings by the dominant multinational (MNE) record companies provides an excellent case study of the complexities surrounding the relationship between copyright law, international trade and competition policy.

**Anti-Competitive Conduct in the Post-Reform Period: A Case Study**

In 1999 the ACCC initiated legal action against Universal Music Australia, Warner Music Australia and Sony Music Entertainment (Australia) alleging conduct that breached Sections 45, 46 and 47 of the TPA. Prior to the commencement of the trial, the ACCC and Sony reached a settlement whereby Sony gave an undertaking to refrain from conduct designed to impede parallel imports and parallel exports.

*The music recording industry*

Record companies are in the business of producing and distributing sound recordings. Performing artists are signed to exclusive contracts, typically encompassing multiple albums (sound recording titles). Record companies own copyright in the sound recordings they produce, and pay royalties to the performing artist(s) and songwriter(s) for each copy of the sound recording reproduced. In this way, a record company has monopoly control over its catalogue of artist specific sound recording titles.
The Australian sound recording market is dominated by four record companies (the ‘majors’), which are subsidiaries of MNEs. Together these record companies share around 90 per cent and 80 per cent of the domestic and international markets respectively. Not surprisingly, international repertoire represents around 85 per cent of all sound recordings sold in Australia. The majors have exclusive territorial licenses for the parent company’s foreign music catalogue. Prior to 1998 music retailers had to deal exclusively with record companies to source artist specific sound recording titles over which they had monopoly distribution rights. This provided the majors with a considerable degree of market power in the wholesale distribution market. This influence over the market for sound recordings was used in an attempt to maintain monopoly control over their respective music catalogues.

The conduct

The removal of the importation right provided the opportunity for the establishment of competing unauthorised distribution channels for sound recordings that divert sound recordings from relatively low-price markets into Australia. These competing distribution channels provide music retailers with alternative sources of supply and introduce intra-title competition and place downward pressure on prices.

Record companies responded by adopting strategies, both domestic and international, that would impede parallel imports. The domestic conduct included the withdrawal of trading terms to music retailers\(^2\), and in some cases the outright cessation of supply, for those deemed to be buying parallel imports from competing distributors or directly engaging in parallel imports themselves. In the view of the ACCC, this adverse treatment was a form of signalling to other music retailers as a deterrent to dealing with competitors.

The international dimension to the actions undertaken by the major record companies included attempts to impede parallel exports from relatively low-price markets, such as Indonesia. This was achieved by putting pressure on Indonesian distributors, via affiliate record companies and distributors in that territory, to not supply Australian retailers or independent wholesale distributors.

Breaches of the Trade Practices Act

In considering the merits of the case, Justice J. Hill found that both Universal and Warner were in breach of Sections 46 and 47 of the TPA (Federal Court of Australia, 2001).

Section 46 prohibits the use of market power that has the effect of eliminating competition or impeding entry into a market. The removal of the importation right provided the opportunity for new entrants into the wholesale distribution market

\(^2\) Trading terms typically offered to retailers included bulk discounts, credit terms, sale-or-return policy and cooperative advertising.
and for local retailers to source products from distribution channels located in foreign territories. The actions undertaken by the record companies were designed to prevent parallel importation of sound recordings by competing distributors and retailers. While the record companies did not have any direct control or influence over competing distributors, the removal of trading terms and, in some cases, the closure of accounts (cessation of supply) for retailers that stocked parallel imports, discouraged retailers from buying relatively low-price substitutes from a competitor. Both companies were found to have breached section 46, having used their market power to prevent entry into the wholesale distribution market.

Section 47 (exclusive dealing) prohibits the imposition of conditions on customers that prevents them from dealing with competitors. By threatening to impose sanctions on retailers found to be directly importing or sourcing parallel imports from a domestic competitor, the record companies attempted to force retailers to deal exclusively with them. Accordingly, they were judged to have breached section 47 of the TPA. In addition, a number of senior record company executives were also judged to be guilty of accessorial liability by knowingly engaging in conduct that was in breach of Sections 46 and 47.

The ACCC also brought charges against the record companies in relation to the overseas conduct in which they attempted to impede parallel exports to Australia. Section 45 of the TPA prohibits arrangements or understandings that impede or substantially lessen competition. This action failed because, while there was documentary evidence relating to correspondence between the Australian and Indonesian affiliates, it could not be proved that the arrangement to refuse supply to Australian competitors was entered into by the Australian companies rather than their Indonesian counterparts.

Conclusions

There remains considerable disagreement over the economic rationale for copyright owner control over the international distribution of copyright product and parallel imports. The music recording industry case study highlights the interplay between copyright law, international trade and competition policy.

The recent Federal Court decision demonstrates that former statutory monopoly distributors of copyright product cannot pursue anti-competitive strategies to maintain their monopoly control in an environment of international exhaustion and parallel imports. Foreign copyright owners need be aware of potential competition policy violations of actions by their domestic subsidiaries. The judgment will no doubt have a signalling effect to monopoly distributors of other copyright products should the Government’s reform agenda continue, with the reintroduction of the Bill to remove parallel import restrictions on all remaining copyright products (except motion pictures).
References


Commonwealth of Australia (1998), Copyright Amendment Act (No. 2) 1998, AGPS, Canberra.

Federal Court Of Australia (2001), ACCC v Universal Music Australia Pty Ltd and Warner Music Australia Pty Ltd, FCA 1800, December.


Price Surveillance Authority (1990), Inquiry into the Prices of Sound Recordings, Australia, Report no. 35.

Senate Legal and Constitutional Legislation Committee (1998), ‘Consideration of Legislation Referred to the Committee: Copyright Amendment Bill (No. 2)’, Parliament of Australia, Canberra.


WIPO (1996), WIPO Copyright Treaty (WCT) 1996.


The author is grateful for helpful comments received from two anonymous referees, and to Dr Malcom Abbott for suggesting the writing of this paper for Agenda.
Mathematicians as Great Economists: John Forbes Nash, Jr.

Matthew J. Ryan

Sylvia Nasar’s recent book, A Beautiful Mind, and a film of the same title about the life of John Nash have piqued public curiosity about this most interesting of characters. However, an important aspect to Nash’s story is neglected by these sources. This is the tale of the rise and fall, and rising again, of his ideas, and their ultimately enormous influence on social science. Economic and political theory are now heavily based upon the ‘rational strategic’ models of human behaviour that Nash pioneered. The impact of this new theory on policy is also apparent. It has altered the conduct of monetary policy and the relationship between central bankers and elected public representatives; utterly transformed the work of the ACCC and other regulatory bodies; informed decisions about whether and how to auction public assets; and revised thinking on the negotiation of international trade and other agreements.

What follows is a selective and, more importantly, an interpreted history of why and when the influence of Nash’s ideas came to be so strongly felt, particularly within economics. I hope that the history is accurate; but the interpretation is of course subjective, and proffered as a stimulus to further debate.

The nub of the debate is, I believe – what makes a mathematician a great economist? As is well known, Nash is a mathematician, and attended but a single course in economics — undergraduate International Trade at Carnegie Tech (now Carnegie Mellon University). The work by which he is best known to economists is contained in three papers — a mere 30 pages — published between 1950 and 1953: Nash (1950b; 1951; 1953). Indeed, his Nobel citation refers primarily to the second of these, which contains the substance of Nash’s PhD thesis. After 1953, Nash’s attention turned to mathematical problems of no obvious interest to economists.

Why should a mathematician, with such a paucity of direct involvement in the enterprise of economics research, have been bestowed (albeit amid unprecedented controversy) with two of economics’ highest honours: the Nobel Prize (1994) and fellowship of the Econometric Society (1990)? The case of Nash’s PhD supervisor, Albert Tucker, provides a useful comparison. In 1951, the very same year in which Nash’s Nobel Prize-winning paper appeared, Tucker and his co-author, Harold Kuhn, published a famous paper on the mathematics of optimisation. The ‘Kuhn-Tucker Theorem’ now appears routinely in economics textbooks, and is known to virtually every graduate student of the subject. But Kuhn and Tucker are no more regarded as part of the economics Pantheon than are

Matthew Ryan is a Lecturer and ARC Senior Research Associate in the School of Economics at the Australian National University.
the inventors of calculus. Just because economists use mathematics, this does not make mathematicians famous contributors to economics. Why has John Nash been singled out?

To understand the answer, we must relate the tale of the turbulent relationship between mathematics and economics. Let us begin by describing the intellectual landscape into which Nash first makes his appearance.

**Setting the Scene**

The year 1948, in which John Nash entered Princeton as a new PhD student, saw the confluence of two powerful forces that would have a lasting impact on economic research. The first was the increasing acceptance of mathematical formalism in economic analysis, culminating in the publication of Paul Samuelson’s *Foundations of Economic Analysis* in 1947.

The second was the emergence of a new field of mathematics, called game theory. John von Neumann’s and Oskar Morgenstern’s *Theory of Games and Economic Behavior* was the first substantive treatise in this new field. Its first edition appeared in 1944, and an important second edition — the first to include the famous expected utility model — in 1947.

As the title of von Neumann’s and Morgenstern’s book suggests, and its first chapter puts beyond doubt, the authors were convinced that game theory held significant promise as a tool for mathematical economic research. The rapid convergence of these two streams of intellectual development therefore seemed imminent.

*Mathematical economics and the ‘neoclassical’ agenda*

The first major economic treatise to make substantial use of mathematics — that is to say, elementary calculus and algebra — was the *Recherches sur les Principes Mathematiques de la Theorie des Richesses* (1838) by Antoine Augustin Cournot. This was, at the time, a rather radical departure from the norm, and the book was not well received, prompting Cournot to publish a non-mathematical version, the *Principes de la Theorie des Richesses*, in 1863. Far from being lauded for introducing new discipline and rigour into academic discourse on political economy, the mathematical method was scorned as irrelevant window-dressing, or worse, as a deluded subversion of social science.

However, by the late 19th century, the tide was beginning to turn. The impressive work of Jevons, Walras, Edgeworth, Pareto, Marshall and others slowly won over the nay-sayers. Acceptance of the utility of mathematical analysis spread amongst the agenda-setters of the discipline. As Irving Fisher (1898:136) observed in reviewing the 1897 English translation of Cournot’s *Principes Mathematiques:

Marshall’s diagrams and formulae were called dangerous, falsely accurate, academic playthings. ... To-day few economists can be found
who regard diagrams as useless curiosities or as waste of valuable page space.

It should be noted that diagrams of this sort are now used to teach first-year undergraduates! The cogniscenti of 1898 may indeed have embraced the mathematical method, but the level of technical analysis was rather low.

The next milestone in the progress of mathematical economics was Samuelson’s *Foundations*, which ‘actively contributed to raising the level of analysis in economic science’. So reads Samuelson’s 1970 Nobel Prize citation. Samuelson (http://www.nobel.se/economics/laureates/1970/samuelson-bio.html) himself put it more pithily, claiming his mathematics saved economists from practising ‘mental gymnastics of a peculiarly depraved type’. His *Foundations* also brought the new mathematical approach to the rank-and-file of graduate students, academics and other practising economists.

However, it is inevitable that the desire to impose the discipline of mathematical formalisation on economics should entail a narrowing of the research agenda. Some topics were better suited than others to expression in precise mathematical terms.

As an example, suppose that one is interested — as economists are — in understanding the functioning of competitive markets. A natural starting place for the analysis is a bargain struck between two traders. However, pre-War economists had virtually given up hope of formalising the process by which bargaining determines the division of surplus; that is, price. This is vividly expressed in the exasperated words of Edgeworth (1881:46): the division, he declared, must necessarily be determined by ‘higgling dodges and designing obstinacy, and other incalculable and often disreputable accidents.’

Sadly, the analysis of ‘higgling dodges’ continues to elude adequate formalisation. What are we to do? If one cannot predict the outcome of individual bargains, how are we to understand complex interacting markets with thousands of traders? Fortunately, the study of markets ‘in the large’ turns out to be easier than that of bargaining ‘in the small’. By studying dense markets, with many buyers and many sellers, individual negotiation skills become irrelevant. A ‘market price’ asserts itself inexorably upon the process of trading.

In these densely populated markets, traders cease to eyeball each other across the negotiating table, and instead operate anonymously, consulting a price posted by the ‘invisible hand’ of market forces, and submitting their trades to a central clearing house. This is the paradigm of the so-called Walrasian market, which quickly came to dominate economic theorising.

All direct human interaction is mired in the vagaries of human psychology and emotion. Resorting to the idealisation of *homo rationalis* does little to improve matters. In an interactive social environment, one’s own best course of action can be determined only once one has formed a belief about how others will act. Since these ‘others’ face a symmetric decision problem, an impasse is reached. Like the bargaining problem, many avenues of obvious economic enquiry were blocked in this manner. With face-to-face human interaction
effectively off the agenda, economics had thus become a rather ‘asocial’ branch of social science.

*Game theory*

Game theory is the mathematical analysis of precisely these sorts of human interactions. It is the science of strategic decision-making. In the abstract, a strategic decision may be defined as one whose consequences are not uniquely determined by the decision-maker’s own actions plus exogenous random forces (‘chance’), but are also affected by the deliberate actions of other people. Such problems are fundamental to any social science.

As its name suggests, the origins of game theory may be found in attempts to devise winning strategies in popular parlour games, such as chess, poker and ‘paper, rock, scissors’. Important contributions to statistics and probability arose from the study of ‘games of chance’; and so too did these apparently frivolous musings upon ‘games of strategy’ lead ultimately to powerful insights into such weighty matters as military tactics, oligopolistic behaviour, evolutionary biology, and politics.

Game theory existed as a branch of mathematics before Nash began writing on the subject, but only just. The French mathematician Emile Borel and the Hungarian wunderkind John von Neumann had written a few papers on the subject in the 1920s, and there were some other early contributors of lesser significance. The key result to emerge from this early literature was von Neumann’s famous Minimax Theorem (1928).

This theorem provided a solution to so-called two-player, zero-sum games. In such games, player payoffs always sum to zero: the winnings of one player must be exactly matched by the losses of the other. The Minimax Theorem suggested an optimal way to play such games.

As a simple example, consider the game of Matching Pennies. The players are designated One and Two, and each player places a penny on the ground simultaneously. If both match (that is, both pennies are ‘heads’ up, or both are ‘tails’ up), Player One keeps both pennies; otherwise both pennies go to Player Two. This is a zero-sum contest, since one player must gain a penny and the other must lose one. The game is illustrated schematically below. The payoffs to any given strategy pair is indicated in the corresponding cell of the matrix, with Player One’s payoff listed first.

<table>
<thead>
<tr>
<th>PLAYER ONE</th>
<th>Heads</th>
<th>Tails</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heads</td>
<td>+1,-1</td>
<td>-1,+1</td>
</tr>
<tr>
<td>Tails</td>
<td>-1,+1</td>
<td>+1,-1</td>
</tr>
</tbody>
</table>

Because each player’s attempt to maximise her own position in a zero-sum contest is simultaneously an attempt to inflict maximal damage upon the other, it
Mathematicians as Great Economists: John Nash 125

is natural to consider the worst-case scenario associated with each strategy (that is, the scenario in which the rival’s strategy inflicts the minimal payoff). For any given strategy, its worst-case scenario payoff is called its security level. A maximin strategy is one that maximises the player’s security level. This is an obvious ‘defensive’ strategy in such games.

Interestingly, as von Neumann and others had recognised, a maximin strategy may require the player to randomise. The maximin strategy for a player in Matching Pennies, for example, is to toss the penny, rather than place it. This gives a 0.5 probability of ‘winning’ no matter what the other player does. Hence, the tossing strategy guarantees a payoff of zero (in expected value). Any other strategy, randomised or otherwise, has a security level that is strictly less than zero. (When players toss their coins, Australian readers may recognise Matching Pennies as a simple version of Two Up.)

The Minimax Theorem demonstrates that if one player adopts a maximin strategy, then the other can do no better than to do likewise. Therefore, maximin strategies are mutual ‘best responses’. They are in a relationship of equilibrium with respect to each other.

It therefore seems compelling that the best advice to give a player in a two-person, zero-sum contest is to play a maximin strategy. Importantly, while players in some games may have many different maximin strategies, the Minimax Theorem implies that it doesn’t matter which is chosen. Each can be expected to perform as well as the other(s). This is sometimes expressed by saying that the strategies are interchangeable.

In 1944, with the publication of von Neumann’s and Oskar Morgenstern’s Theory of Games and Economic Behavior (henceforth TGEB), the scattered contributions to game theory were synthesised and extended. It also set an agenda that took considerable hold of the discipline for years to come.

For our purposes, two features of this book are noteworthy. First, it contains essentially no new important game-theoretic results. Its theoretical centrepiece remains the Minimax Theorem. In particular, the authors’ analysis of non-zero-sum games, and games with more than two players, produced a lot of interesting ideas, but no firm conclusions. A ‘solution’ for such games remained elusive.

Second, Morgenstern, an economist at Princeton, had perceived the potential value of game theory to economics, and convinced von Neumann to pitch the book in this direction. Indeed, TGEB was very favourably reviewed by leading economic theorists such as Leonid Hurwicz (1945), Jacob Marschak (1946) and Richard Stone (1948).

Unfortunately, most economic problems are not zero-sum in nature. For example, the bargaining problem involves the potential for mutual advantage through trade, and with respect to bargaining, TGEB was unable to improve on the dispiriting conclusion of Edgeworth. Morgenstern was aware of this imperfect match between the tools that TGEB was able to provide and the needs of economists. Nevertheless, in the section on ‘Necessary Limitations of the Objectives’ he recalls from the history of the physical sciences that ‘in the study of problems which were modest as compared with the ultimate aims, methods were
developed which could be extended further and further’ (von Neumann and Morgenstern (1947:6)).

And so, the stage is set...

The year 1948 dawns, and Mr Nash goes to Princeton, the birthplace of GEB. Economists are increasingly receptive to sophisticated mathematical ideas, and Morgenstern has alerted them to the potential value of game theory.

By the time Nash has completed his PhD in 1950, he will have provided precisely the extensions to GEB that economists sought, including a solution to the famous bargaining problem. All seems in readiness, then, for economics to embrace game theory. Instead, the basic ideas of game theory and its powerful applications to economics were largely ignored for the next twenty-five years.

Two Important Results of John Nash

Let us first examine Nash’s two key contributions — Nash equilibrium and the Nash bargaining solution — in more detail. The first of these is a contribution to the theory of non-cooperative games: those in which the players are unable to write enforceable contracts. The bargaining problem, by contrast, is a cooperative game, since the bargain may be written down and enforced through an external mechanism (that is, the courts). Indeed, this important distinction between cooperative and non-cooperative games is another of Nash’s contributions.

Nash equilibrium

Recall that the appeal of the maximin solution to zero-sum games arises from two features: maximin strategies offer a natural defensive posture; and a situation of ‘equilibrium’ results when players adopt such strategies. Nash observed that the first property does not survive outside the zero-sum context, but showed that the second alone is sufficient to extend the solution beyond it. A Nash (1950a; 1951) equilibrium is a collection of strategies — one for each player — that are mutual best responses. For zero-sum games, only maximin strategies may be played in a Nash equilibrium, but Nash proved that equilibrium strategies exist for any non-cooperative game, including games that are not zero-sum, and those with more than two players. This is Nash’s famous equilibrium existence result.

Nash equilibrium also conveyed a powerful contrarian message to economists enamoured of the efficiency of Walrasian markets. Consider the following example, cited in Milnor (1995). Some readers may recognise it as a variation on the classic Prisoners’ Dilemma game, which is attributed to Albert Tucker but first appeared in print in Nash (1951).

A group of 20 people is going to dinner, and each must choose between a $10 meal and a $20 meal. Each values the expensive meal only $5 more than the cheaper one, so if each pays individually, all diners choose
the $10 meal. But if the bill is split, the meal choice becomes a strategic problem, and the only Nash equilibrium is for each to choose the $20 meal: any individual who opted for the $10 meal would forego $5 worth of additional value which they might have secured for only 50 cents.

When the bill is split, each diner obtains some ‘market power’ over the ‘price’ paid by the other diners for their meals. The individually rational exercise of this market power leads to a socially inefficient outcome. Everyone is made worse off.

Of course, economists in 1950 were already aware of the efficiency problems with imperfectly competitive markets, but Nash’s equilibrium provided a powerful unifying principle with which to understand and generalise these phenomena.

*The Nash bargaining solution*

In bargaining problems, it is assumed that the capacity to sign binding contracts eliminates efficiency problems. If an enforcement mechanism were available, Milnor’s diners would readily have signed contracts committing themselves to the cheaper meal, whether or not the bill was to be split. In bargaining, the problem is not to ensure that maximal surplus is achieved, but to determine how it will be shared.

Abstractly conceived, a two-person bargaining problem consists of a set of feasible contracts, with associated payoffs to each bargainer. As Edgeworth, von Neumann and Morgenstern, and others had observed, it is reasonable to suppose that the chosen contract will satisfy:

- **Efficiency**: No other feasible contract offers both bargainers a higher payoff.

However, this requirement typically leaves a large number of possibilities. Choice amongst them, it was felt, could only be determined by understanding the complex psychology of bargaining. Nash (1950b; 1953) demurred. The process by which the ultimate bargain is negotiated may be complex, he agreed, but elementary principles of rationality dictate some additional properties of its outcome.

- **Symmetry**: If the bargaining problem is symmetric (that is, if for every feasible contract, there exists another which swaps the bargainers’ payoffs), then the chosen contract should award each bargainer the same payoff.
- **Independence of Irrelevant Alternatives**: If bargaining reaches a particular outcome, and we discard some of the unused contracts and ask the parties to renegotiate over the smaller set of contracts, they should opt for the same outcome as before.

While these principles are not without their critics, particularly the latter, the remarkable thing is that they should be sufficient to determine a unique outcome.
to any bargaining problem. Using an argument of almost magical simplicity, Nash showed that these three principles can only be satisfied if the bargainers choose the contract that maximises the product of their payoffs. The bargaining problem had been brilliantly solved.

**Denouement delayed**

Precisely why Nash’s results failed to ignite the ardour of economists at the time remains unclear. However, at least two factors appear to have been important.

First, mathematicians held the ‘line’ of TGEB, and rejected Nash’s equilibrium as an inadequate solution for general non-cooperative games. This bad press from the mathematics community, and the consequent absence of mathematicians actively developing Nash’s ideas, must certainly have been discouraging to potential end-users.

Second, economists proved reluctant to expand their horizons beyond the entrenched neoclassical agenda. Appropriate new questions simply did not get asked. To the extent that game-theoretic ideas infiltrated economics, they were the ideas that economists found useful in understanding Walrasian markets.

*The normative agenda in game theory*

When Nash described his equilibrium existence result to von Neumann, the latter cut Nash’s explanation short with the dismissive words (Nasar, 1998:94): ‘That’s trivial, you know. That’s just a fixed point theorem.’

Other mathematicians working on game theory were similarly unimpressed by Nash’s ideas. This is evident in the influential *Games and Decisions* by R. Duncan Luce and Howard Raiffa (1957) — a highly successful attempt to make game theory accessible to a wider audience of social scientists. After reviewing some of its properties, Luce and Raiffa (p. 104) find Nash equilibrium wanting, and conclude that ‘a unified theory for all possible non-cooperative non-zero-sum games does not seem possible’.

This judgement sounds almost shocking to modern ears. Nash equilibrium is now accepted without question as the way to ‘solve’ games. Why did mathematicians reject it?

Let us examine von Neumann’s charge of ‘triviality’ first. With minimal reading between the lines of history, it is entirely plausible that von Neumann was either aware of the possibility of proving Nash’s result, or capable of reaching such a conclusion within a few short logical steps. The logic of equilibrium was familiar to him from his zero-sum analysis, as we have already observed. Moreover, he was also familiar with the basic mathematical tool — fixed point theory — that Nash used to extend this equilibrium logic to general non-cooperative games. Indeed, for a paper on economic growth written in 1932 — von Neumann’s one other foray into economics — von Neumann (1945-6) proved a fixed point theorem essentially equivalent to the one applied in Nash (1950a), an
early version of Nash’s equilibrium existence result\(^1\). In his growth paper, von Neumann (1945-46:5.) even observes that the Minimax Theorem follows as a direct corollary of this fixed point theorem. Thus, von Neumann was aware of the logic of equilibrium, acquainted with fixed point theory, and aware of the utility of the latter in connection with the former. Nash’s result was surely within his grasp.

It is not surprising, therefore, that von Neumann saw Nash’s result before Nash had finished his explanation, and saw little novelty in it. One is therefore lead to the question: Why had von Neumann not already written down the extension of equilibrium to general non-cooperative games? The answer is evident in the lukewarm reception to the idea expressed by Luce, Raiffa and others in the game theory community. Game theorists simply did not think Nash equilibrium a viable ‘solution’ to non-zero-sum games. They felt it to be conceptually flawed.

The reason for this, it seems clear, was the normative orientation of mathematical game theory. True to its origins, game theory in the 1950s still sought to advise players on how best to make their strategic decisions. Nash equilibrium failed to generate unambiguous advice, since many non-zero-sum games possess a multiplicity of equilibria that are not interchangeable. The best advice to give one player depends upon the advice being received by the other, even if both are being advised according to Nash’s theory, and so we arrive at yet another strategic impasse.

Social scientists, of course, are not concerned to advise people how the ought to behave, but rather to understand how they do behave. They required a positive theory of games. Nash equilibrium served this descriptive purpose perfectly. Indeed, even Luce and Raiffa (1957:105) tempered their criticism of Nash with the observation that:

> Even if we were to reject equilibrium as a normative theory for non-cooperative games ... it may still be that the notion is relevant as a description of behavior.

What evidence is there that Nash interpreted his own work in a positive spirit? On the one hand, his 1951 paper reserves the term ‘solution’ for cases in which a game’s Nash equilibria are interchangeable. This suggests a normative inclination. However, one may also argue that Nash was merely adopted a usage of the term ‘solution’ that was consistent with its normative origins in the earlier work on zero-sum games. In other words, he may have seen his equilibrium as a satisfying descriptive analysis for all games, but an adequate normative solution only for games in which the additional requirement of interchangeability is satisfied.

Indeed, there is compelling evidence elsewhere that Nash did seek to motivate his equilibrium as a positive concept. Much of this evidence was

---

\(^1\) von Neumann (1945-6) is an English translation of von Neumann’s paper, originally published in German in 1938.
obscured from public view because of the omission of a section on ‘Motivation and Interpretation’ from the published version of Nash’s PhD thesis. In this section, Nash (cited in van Damme and Weibull (1995, p.19, emphasis added)) says that he wishes to answer the question: ‘(W)hat would be a rational prediction of the behavior to be expected of rational players in the game in question?’.

He also goes on to propose what he calls a ‘mass action’ interpretation of equilibrium. He suggests that equilibria may come to be played through an evolutionary process by which a population of players are randomly matched and re-matched in the same contest. Their strategy choices evolve, and will only stop evolving once an equilibrium has been reached, so that no deviating ‘mutant’ can obtain a superior outcome. This interpretation was independently re-invented in the 1970s, and evolutionary game theory is now a thriving branch of the discipline.

In the ‘mass action’ story, the strategies of all players evolve together. There is no need of interchangeability, since strategies are jointly determined. If multiple equilibria exist, this only reflects the inability of Nash’s logic to generate unique predictions. The players themselves know perfectly well which equilibrium they are playing; only we, as outside observers, are in doubt.

One also sees a methodological unity between Nash equilibrium, under the ‘mass action’ interpretation, and his bargaining solution. In both cases, Nash uses elementary principles to narrow down the plausible outcomes of an underlying process of considerable complexity. This methodology is now unquestioningly accepted and powerfully applied by economists and other social scientists. However, it was certainly not embraced immediately.

**The apotheosis of Walrasian market theory**

Why it was not so embraced remains an intriguing unsolved mystery. Certainly, leading economists had more than enough mathematical facility to read and interpret Nash for themselves. Indeed, Nash’s papers were read, and their ideas syphoned off by economists. However, in one of the great ironies of our story, the ideas were applied to understand the non-strategic behaviour of Walrasian markets, not strategic interaction.

Kenneth Arrow and Gerard Debreu were directly inspired by Nash’s fixed-point arguments (Debreu, 1952; Arrow and Debreu, 1954). By adapting these arguments, they were able to show that, under fairly weak assumptions, there will always exist a set of prices that balance supply and demand simultaneously on a system of Walrasian markets. This is known as a Walrasian equilibrium.

Even the Minimax Theorem exerted an indirect influence on economic theory at this time. Tjalling Koopman’s activity analysis, for which he won a Nobel Prize in 1975, is an application of duality theory in linear programming, which is

---

2 Interestingly, Debreu’s paper was communicated to the National Academy by John von Neumann.
Mathematicians as Great Economists: John Nash 131

itself a corollary of the Minimax Theorem. Economists were not only out of practice at asking the sorts of questions that game-theoretic tools were useful in answering, the usefulness of the mathematics of games to the neoclassical agenda contributed to the temporary demise of game theory itself within economics. Indeed, this traditional agenda acquired a new momentum that sustained it into the 1970s.

The Game-theoretic Revolution

As Aumann (1985:43) states:

The Nash equilibrium is the embodiment of the idea that economic agents are rational; that they simultaneously act to maximise their utility. If there is any idea that can be considered the driving force of economic theory, that is it.

By the mid-1970s economics was ready for a new paradigm. The heyday of research into Walrasian markets was over, and graduate students were looking for new topics on which to write their dissertations. A game-theoretic revolution in economics was about to take place.

Of course, economists had never entirely abandoned game theory, or Nash’s ideas. Both had persisted at the fringes of the discipline, sustained by the work of theorists such as Robert Aumann, John Harsanyi, Lloyd Shapley and Martin Shubik. It is also true, as I have attempted to convey, that the ideas of Nash were so compellingly useful to economists, that it was only ever a question of ‘when’ they would be adopted, not ‘if’. Nevertheless, an explanation is required for their suddenly being thrust centre stage at this particular juncture.

Definitive answers are elusive, but some influential factors are evident. The first is the aforementioned loss of steam from the research program into Walrasian markets. Second, political scientists have suggested that game theory may have re-entered economics via politics, where it had spawned the important new ‘positive political theory’ school, lead by William Riker, in the 1960s (Amadae and Bueno de Mesquita, 1999). Indeed, the use of game theory in political science has seen a growing together of economics and politics. A new political economy, formalised along game-theoretic lines, has emerged. Theories of voting, lobbying and other political phenomena are now routinely taught in economics programmes.

Finally, there was the important catalyst provided by Nash’s two Nobel co-laureates: Reinhard Selten and John Harsanyi. Selten pioneered the literature on so-called refinements of Nash equilibrium. By tightening the predictive power of equilibrium, the refined solutions greatly enhanced the attractiveness of game theoretic tools for applied work. Harsanyi enhanced Nash equilibrium in the opposite direction, by expanding its scope of application. Nash, like the game theorists before him, had assumed that all players knew the payoff consequences for all rivals of any given outcome of the game. This complete information assumption is frequently unrealistic. For example, in the context of an auction, it
requires that each bidder know every other bidder’s valuation of the object for sale. Harsanyi indicated how the complete information assumption might be dispensed with using an elaboration of the Nash logic known as a Bayesian Nash equilibrium. A vast new array of interesting problems could now be successfully tackled.

With respect to the contributions of Selten and Harsanyi, however, let me not give the false impression that Nash’s ideas were inadequate in their own right. It was not that Selten’s and Harsanyi’s embellishments were necessary to make Nash equilibrium serviceable. The new ideas re-ignited interest in game theory, but they did not render Nash equilibrium obsolete. On the contrary, as Aumann (1985:48) observes: ‘Nash equilibrium is without doubt the most ‘successful’ — that is, widely used and applied — solution concept in game theory.’ Indeed, most applied work on problems of complete information employ nothing more than simple versions of the Selten refinement that combine Nash equilibrium with the even older idea of ‘backward induction’. This combination is already evident in Nash’s own work: see Nash (1953).

**Game Theory and the Modern Agenda in Economics**

To conclude, let me return to the question: what distinguishes the contributions of Nash from those of his supervisor, Tucker? Unlike any other single idea, Nash equilibrium has transformed the agenda in economics, and in social science more broadly. While the Kuhn-Tucker Theorem allowed economists to do what they had already been doing more effectively, Nash’s equilibrium opened entirely new avenues of research, and substantially altered both the scope and the focus of economics. In particular, the credit for this must go to Nash himself, rather than to the collective enterprise of game theory. Nash was the first to see clearly how game theory might be used in a positive mode, and is one of its most accomplished exponents. Reading the game theory literature of the 1950s, it is striking how Nash stands out as the most ‘modern’, the most in tune with contemporary thinking about strategic analysis in social science. This is indeed no accident.

It scarcely needs to be emphasised that game theory is now ubiquitous within economic research, from industrial organisation, whose existence one could scarcely imagine without it, to strategic trade theory, law and economics, and even macroeconomics. Indeed, game theory itself now exists as a field within economics, entirely separate from the kindred branch of mathematics.

The analysis of markets now focuses heavily on the ‘visible hand’ of the strangely mis-named ‘imperfect competition’, rather than the ‘invisible hand’ of the shadowy Walrasian mechanism. To the extent that Walrasian markets are the object of continued theoretical interest, this is directed at providing strategic foundations in terms of individual bargaining. Nash’s bargaining solution is the key ingredient in this endeavour.

Most importantly, perhaps, is the way in which game theory has transformed the ‘dismal science’ into a truly ‘social’ science, and reconnected it with its
brethren: political theory, sociology, law, anthropology, and even biology. Social and economic institutions are now viewed as mechanisms that determine the rules of the games that people play. Normative questions of institutional design (so-called mechanism design or implementation theory) may be addressed in an intellectually satisfying manner. Dialogue on these issues of common interest is now possible between economists, political scientists and other social philosophers. This, ultimately, may be the most significant legacy of Nash (Myerson, 1996:273-4):

Methodological limitations no longer deter us from recognising the essential interconnections between economic, social, and political institutions in economic development. ... Today, the original scope of political economy has been restored to economic theorists, because of the general methodology that John Nash introduced.

References


Luce, R. and H. Raiffa (1957), Games and Decisions, John Wiley and Sons, New York.


This paper is based on a Toyota/ANU Public Lecture, 21 March 2002. My thanks to the School of Mathematical Sciences at the ANU for inviting me to write and speak on this topic. Thanks also to Gerry Mackie, Guillaume Rocheteau, two anonymous referees, and seminar participants in the School of Economics at the ANU, for helpful comments on an earlier draft.
No-Fault Public Liability Insurance: Evidence from New Zealand

Bronwyn Howell, Judy Kavanagh and Lisa Marriott

Public liability insurance indemnifies an organisation against claims made by individuals for illness, injury, loss or damage to property. The spiralling cost of public liability insurance has been a recurring problem in the United States, England, Australia, and elsewhere for well over a decade (for example, Danzon, 1990; Kessler, 1999).

The cost of public liability insurance is a product of the frequency of the claims made, coupled with the size of the damages awards, plus transaction costs — the administrative costs and profits of the insurance companies, as well as the costs of justice (Goodman and Thanhauer, 2001). Recent events such as the September 11 attack in New York have highlighted the consequences of increases in both the probability of loss and the size of potential claims. As insurance companies raise premiums to cover the increased risk and the size of loss, the potential for a missing market for public liability insurance has increased. Moreover, the consequences of missing markets can extend far beyond those seeking to insure themselves against liability, and their prospective insurers. For example, societal benefits are lost when community groups cannot hold social functions because the cost of public liability insurance for hiring a venue becomes prohibitive, and where patients incur health-related costs because doctors, unable to procure medical liability insurance from a private market, refuse to treat them.

Governments, as insurers of last resort, are increasingly being called upon to address these missing markets, either by underwriting some public liabilities themselves or legislating to impose compulsory no-fault insurance schemes, offered by either private sector or government-owned insurance companies. No-fault schemes have attracted significant interest and have been proposed as an instrument potentially capable of simultaneously ‘solving’ two important insurance-based problems: providing products in the missing markets; and reducing the sharply escalating costs of tort actions that are perceived in some cases to be giving rise to missing markets. For example, Carroll et al. (1991) report that the transaction costs to private parties in tort actions involving bodily injury account for 33 per cent of compensation expenditures and would be even higher if government spending on courts was included in the total costs. Furthermore, the costs may not always be direct. Cooter and Ulen (1986) argue that potential injurers may take socially excessive precautions to decrease the costs.
of expected liability, especially when there is a high probability of mistakenly being found liable.

While governments have a role when there is a missing market, policy intervention must be informed by analysis based upon the economics of insurance and relevant experience. The ability to pool risk (Milgrom and Roberts, 1992), combined with the power to compel (Horn, 1995), suggests that in some circumstances governments may be more efficient in offering insurance than the private sector. Centralised schemes with scale economies (whether provided by government or the private sector) may also offer transaction cost advantages and greater certainty of being able to deliver compensation when an adverse event occurs (Cummins, Phillips and Weiss, 2001). However, incentives and their effect upon participants’ behaviour and hence the costs and probabilities of loss, should also be considered. Evidence from existing no-fault schemes can provide an empirical ‘litmus test’ of the impact of no-fault schemes in practice.

New Zealand’s experience with its no-fault, government-underwritten accident compensation insurance, which provides compensation for personal injury but carries with it the waiver of the right to sue, provides one such test. A reliable counterfactual is difficult to determine. But evidence from this scheme leads us to conclude that, while providing certainty of compensation and apparent (though unverifiable) reductions in transaction costs, the absence of both a tort-based cost recovery process and risk-based premiums means that the level of care taken is less than socially optimal, resulting in total costs that rival or may even exceed those of tort-based systems.

This is not to say that there is no role for no-fault insurance systems. Rather, we suggest that judiciously applied, there may be a role for such schemes where the need for certainty of compensation is high, but the opportunities for individuals to manipulate the probability of an adverse event occurring is low (for example, natural disasters). However, we caution against discarding the right to sue and losing the ability to risk-rate individuals, when the cost of schemes can be manifestly affected by moral hazard actions. In other words we caution against giving up the incentives embodied in tort law and private insurance.

**Liability and Insurance**

Liability law, or tort law, allocates the cost of accidents among individuals. Tort law has two principal goals (Kessler, 1999:1): to compensate parties suffering loss; and to provide potential loss-causers with incentives to avoid the acts that incur loss.

Tort law is efficient if the resulting level of adverse events is socially optimal — that is, when the social costs of preventing loss are less than the costs created by the adverse event. The deterrence goal of tort law is to induce individuals to internalise the negative externality of careless behaviour by making them liable for the losses they cause. If taking precautions is costly, and individuals are not required to pay for the losses that they cause, then they will take fewer precautions than is optimal and more adverse events will occur.
By purchasing insurance, individuals and firms can indemnify themselves against the financial effects of adverse events that may be partially or fully caused by themselves or others. Insurance thus ensures individuals are compensated, even though an individual may be responsible for causing the loss and even though a tort process may be required to extract the costs of the loss from that individual. However, if an individual has purchased insurance to protect himself from bearing the full costs of loss incurred by others as a result of his actions (either deliberate or negligent), he will not bear the full cost of his actions and may then find it personally more efficient to take less care to prevent loss than if the insurance had not been purchased. This is the classic moral hazard problem of insurance (Milgrom and Roberts, 1992). Insurance companies generally seek to recover the additional costs of less care by varying individuals’ premiums by a factor representing the individual’s likelihood of acting without due care (that is, by charging premiums that are risk-adjusted).

By combining the deterrence effect of possible tort action with individual risk-adjusted insurance premiums, both the costs of loss and the optimal level of care may be extracted from potential loss-causers (that is, internalising the costs of harm) without necessarily having to resort to court action. However, the record of tort and insurance-based systems in delivering both compensation and deterrence has been poor (Huber, 1988). Evidence from the United States, for example, points to high transaction costs, distorted incentives and liability pressure leading to over-investment in precautionary practices (Kessler, 1999:2).

No-fault systems have been proposed as a means of reforming the tort-based system. No-fault systems remove the need to prove cause from the compensation equation and thereby reduce the transaction costs of the process (Danzon, 1990:4). Compensation is generally paid according to a prescribed schedule if the loss-sufferer satisfies the scheme administrators that loss has been incurred, regardless of any fault by the alleged loss-causer. Such schemes are typically funded out of levies on potential loss-causers or from general taxation and are accompanied by some limitations on the loss-sufferer’s rights to seek redress from the loss-causer through tort actions.

The principal weakness of no-fault schemes is the difficulty of ensuring that the socially optimal amount of care is taken by potential loss-causers, as the links between their potential to cause loss and the costs of their actions are severed. Consequently, most no-fault schemes are accompanied by additional administrative systems, not present in typical tort systems, to monitor the behaviour of potential loss-causers in order to preserve their incentives for appropriate loss-avoidance (Kessler, 1999:13). The efficiency-improvement criterion is met if the transaction costs saved from the tort process are greater than the additional monitoring costs incurred in the no-fault system for the same level of compensation paid to individuals for the same number of adverse events.

No-fault insurance schemes are common where the probability of a loss occurring is very uncertain. These include ‘acts of God’ (a sudden and unavoidable occurrence caused by natural forces, such as a flood or earthquake), where there is no party obviously at fault from whom partial cost recovery can be
sought. The low probability of these events occurring, the unpredictability of the size of the loss and the exposure to large number of claims arising from a single event, mean that there is often a failure of private insurance companies to indemnify such losses. Further, as no individual has the ability to ‘cause’ the event, there is no need for incentives to prevent the loss-causing event and monitoring costs are low.

However, it is noted that it is still necessary to take steps to incentivise against loss-increasing behaviour by the individual whose losses are indemnified (for example inflating the value of the claim, or choosing to live in an earthquake-prone area) to obviate the moral hazard problem (Milgrom and Roberts, 1992). In these cases, compensation can be achieved more efficiently by a no-fault scheme, provided the levy paid by those covered is proportional to the size of the claim they will make in the event of a loss occurring.

No-fault schemes thus appear to be a viable way of compensating losses from terrorist activities. In such instances, access to compensation via the criminal and tort process is likely to be costly (if it is even possible), very few individuals have the ability to affect the probability of such terrorist-related losses occurring, and incentives to internalise the costs and engage in preventative action are probably futile. Indeed, such schemes are common during wartime. New Zealand’s Earthquake Commission (EQC) was originally created as the Earthquake and War Damage Commission with compensation of such losses being a key motive.

No-fault schemes have, however, been increasingly suggested as solutions to market failure in other areas such as worker injury and public liability compensation and medical practitioner liability. The primary motivations in almost all cases appear to be to ensure certainty of compensation and reduction in transaction costs relative to the court process (Danzon, 1990:4-5).

However, Danzon notes that, while no-fault systems reduce transparent costs, they ‘also generate hidden costs which are much harder to quantify but potentially constitute a much greater real economic cost’. In particular, she stresses that (p.4)

by eliminating liability and investigation into the causes of injuries, (such schemes) reduce expenditures on litigation but at the same time eliminate the information base and the incentives necessary for loss prevention and risk management.

Hence, the losses from the elimination of the tort system go further than just removal of incentives to minimise loss. They also remove the information base from which monitoring activities can be designed and upon which education to prevent future loss relies. Both of these are vital factors in a system that is heavily dependent on overt monitoring to achieve a socially optimal outcome.

In order to overcome these weaknesses, some no-fault schemes attempt to combine the certainty of compensation with experience rating (that is, setting levies to reflect historic loss-causing behaviour by individuals). They also use limited tort action to recover exemplary damages over and above the
compensation schedule in some circumstances, thereby attempting to reintroduce ex-ante incentives to take care (Kessler, 1999;13; Kneiser and Leeth, 1991, cited in Mears and Chapple, 1996:34).

It would appear, then, that no-fault insurance could provide an efficient solution to market failure in two specific circumstances:

- where there is extremely limited ability to manipulate the probability of an adverse event occurring, any moral hazard behaviour influencing the size of the claim is independent of the probability of the adverse event occurring, and the moral hazard can be internalised in the contract between insurer and insured; or

- where tort and criminal laws can act as a adjunct to the no-fault conditions to limit moral hazard behaviour that cannot be internalised in the insurance contract.

Natural hazard insurance, such as New Zealand’s EQC, provides a good example of the first type. Compulsory third party motor vehicle insurance provides an example of the second type. Compensation for loss due to a vehicle accident is assured for the vehicle owner irrespective of who caused the accident. However, the ability for the wronged party to seek punitive damages or criminal penalties (in addition to compensation) for negligent or criminal driver behaviour by other parties, which increase the probability of an accident occurring, acts as an incentive for those parties to act more carefully and thereby reduces the likelihood of road accidents.

No-fault insurance in the absence of tort or criminal law sanctions may, at its ultimate, lead to no ex-ante care whatsoever and the maximum total loss. Monitoring can mitigate this loss. However, this increases transaction costs and will only increase efficiency if the costs of monitoring are less than the reduction in the total cost of loss. Incentives from tort law, on the other hand, will increase efficiency if the penalties exacted are sufficiently high to induce a reduction in moral hazard actions, and hence the total loss, nett of the transaction costs of the tort action (Cutler and Reber, 1998). The choice of regime, therefore, requires a careful analysis of the trade-off of incentive effects and transaction costs against the total cost of loss.

Background to the New Zealand ACC Scheme

New Zealand has operated a no-fault workers compensation scheme for over 100 years, with the initial Worker’s Compensation Act lasting from 1900 until 1974. The initial scheme provided compensation to employees injured in the workplace, but did not cover accidents that occurred outside the employment environment. In 1967, the Report of the Royal Commission of Inquiry into Compensation for Personal Injury 1967 (named the Woodhouse Report, after its Chairman, the Hon. Justice Woodhouse) reported that New Zealanders were experiencing over one hundred thousand industrial accidents per annum. At the time of the Report,
compensation could be sought from three different avenues — the common law of
tort, with an action through the courts; through the aforementioned Workers
Compensation Act, provided the injury occurred at work; or through the Social
Security Act. The Woodhouse Report identified these methods as ‘a fragmented
and capricious response to a social problem, which cries out for co-ordinated and
comprehensive treatment’, concluding ‘no economic reason justifies it’.

The Woodhouse report recommended a no-fault compensation scheme
designed to provide comprehensive insurance coverage for the whole population
‘in respect of industrial, highway and domestic accidents of every kind’ and
certainty over entitlements for claimants. Further, it was estimated that the
recommended scheme could be serviced by ‘much the same amounts in total as
are absorbed at present by various systems operating independently, and which
have a duplication of expense and assistance’. The report is notable for its
recognition of the relevance and importance of transaction costs in the economy.
These costs were perceived to be high due to the presence of 62 insurance
companies creating significant duplication of costs. The authors of the
Woodhouse Report were certain that administrative costs could be reduced to
approximately 10 per cent of existing levels if a single authority handled the
scheme. The Report echoed a growing economics literature (Ronald Coase, John
Commons and others) suggesting that there were potential gains from adopting an
organisational form that would minimise transaction costs.

The Woodhouse report proposed a universal, compulsory, no-fault,
government-administered system for compensation of personal injury to cover all
motor vehicle injuries and all injuries to earners whether occurring at work or not.
The comprehensive nature of coverage stemmed from recognition that workers
do not change their status at 5.00pm, and if injured on the highway or at
home they are the same men, and their needs and the country’s needs of
them are unchanged (Woodhouse Report, s6).

The implicit intention was to remove or reduce the problems inherent in the
Worker’s Compensation Scheme, which included:

- a power imbalance between workers and employers leaving employees
  unable to judge the potential for personal injury risk in the workplace (and
  hence individual risk-bearing);
- the need to prove negligence and the lack of resources to enable workers to
  pursue redress via court actions — less than one per cent of claims made for
  compensation as a result of industrial accident were successful. Moreover the
  expected standard of care was determined objectively, irrespective of the
  defendant’s capacity to guard against risk, and there was no principle of
  proportionality to link the extent of the defendant’s liability to the degree of
  culpability (Todd, 2000);
- the impossibility to predict with any assurance the outcome of a damages
  action — the Woodhouse Report acknowledges that the risks of litigation, the
difficulties of proof, the ability of advocates, the reactions of juries and mere chance itself, turned the system into a lottery;

- long delays and significant cost of the investigations and trial; and
- inconsistencies between eligibility and levels of proof leaving a large proportion of the population ineligible for recompense for injuries that were not caused by their own actions. The Workers’ Compensation Act provided a limited form of compensation, which required proof of fault, while the general system of social security required a means test to be met. The common law action was available to the whole population, but limited resources severely restricted this avenue for many injured individuals.

The Accident Compensation scheme (ACC) was adopted in 1974. It was intended to provide quick and assured compensation from all injuries caused through accidents, with no requirement for proof of fault. The basis for compensation was ‘personal injury by accident’, irrespective of cause. Concomitant with the introduction of the scheme was the removal of the right to sue for compensatory damages related to personal injury. While penalties and sanctions remained in place to punish criminal actions leading to personal injury (for example, dangerous driving), the ability to undertake a tort action was eliminated, except in rare circumstances.

Despite a number of minor modifications to the scheme (including those in the period between 1999 and 2000, when the government monopoly on coverage was temporarily removed and six independent insurance companies, one of which was government-owned, underwrote the scheme) the right to sue for negligence has never been restored. And, apart from the brief period between 1999 and 2000, the no-fault aspect of the scheme has remained intact. There remains little tangible link between cause of accidents and individual liability, and the absence of the potential for tort action means that, beyond the setting of broad industry and activity-based levies, there are no direct individual or employer-based incentives to take due care.

Although it is legally feasible for experience rating of employers to be invoked to recover higher costs and to provide rebates to ‘safe’ employers, limitations in the size of eligible firms mean that only some 5,500 of New Zealand’s 300,000 businesses (fewer than two per cent) face the potential of such incentives (Mears and Chapple, 1996:61). Any propensity for risk-bearing that is outside the ‘average’ level embodied in the levies charged to the vast majority of New Zealand’s employers cannot be adjusted for by individual premiums or incentives. Furthermore, while the previous Workers Compensation Act provided cover only for employees, ACC provides cover for all New Zealanders, yet only employers face any possibility of levies being risk-rated. No such provisions apply to vehicle owners or non-earners whose levies comprised more than 43 per cent of premium income in 2001.

The result is risk-seeking actors having opportunities to engage in moral hazard actions that increase the cost of loss, while risk-averse actors pay an inefficiently high premium which does not recognise that their extra levels of care
lower the probability of loss occurring, and hence total costs. Further, the absence of the right to sue means that there is no tort-based incentive to curb the behaviour of risk-takers that may be negligent, but not criminal. Therefore, the overall efficiency of the scheme is almost entirely dependent on external monitoring and enforcement of participant behaviour. If the costs of monitoring and enforcement are lower than the potential cost of moral hazard actions that monitoring precludes, and result in a lower total cost of claims than an equivalent no-fault scheme backed with tort action, this method would be relatively more efficient for the same level of compensation paid to victims.

Performance of the New Zealand ACC Scheme

If the ACC scheme has been successful, we would expect it would have delivered:

- greater certainty of compensation for injured individuals;
- reduced transaction costs overall; and
- if the trade-off between decreased incentives and increased monitoring has been accurately and efficiently determined, no observable differences in the rate of accident occurrence for the same level of compensation relative to the old Workers Compensation Scheme.

Compensation certainty

Danzon (1990) provides a comprehensive independent review of the New Zealand scheme over its initial years. While changes have been made to the benefit schedule and levying process since then, and the period between 1999 and 2000 resulted in a short discontinuity in relation to risk-rating of employers, the key principles of no-fault and the waiver of the right to sue have remained intact. Thus, Danzon’s evidence remains relevant. She notes (p. 17) that reporting rates for injury rose from 1.5 per 100 population in 1971 to 1.7 per 100 in 1984, indicating that introduction of the scheme resulted in more individuals registering injuries, and hence seeking compensation. Total expenditure also rose from 0.62 per cent of GDP in 1980 to 1.01 per cent in 1986, with total expenditure in current dollars increasing 25.8 per cent in current dollars (average annual compound growth rate of 10.1 per cent) between 1975 and 1989 (p. 18).

Further, Danzon notes that this figure is an underestimation as it does not include the co-payments made by employers covering the costs of lost wages in the first week following injury (about one-half of all lost-time work accidents are estimated to fall into this category). When adjusted for inflation, the level of expenditure in 2000 was approximately 12 times higher than the level of expenditure in 1974 (expenditure of NZ$32.9 million in 1975, equates to NZ$100.6 million in 2000 when adjusted for inflation — the actual expenditure for the latter year was NZ$1.2 billion). Even allowing for the greater numbers of individuals covered by the scheme, this implies a real increase in compensation
has occurred, which Danzon attributes to a significant increase in the number of small claims for minor injury receiving compensation.

Transaction costs

While it is difficult to identify a true counterfactual to the current system, there is strong evidence that the New Zealand no-fault system has reduced transparent transaction costs.

The Woodhouse Report implied that the transaction costs of a single central monitoring and enforcing provider would be more efficient than individual action under the Worker’s Compensation scheme. Centralising collection and distribution of funds from 62 individual insurance companies into a single entity promised a reduction of transaction costs (estimated to equal 40 per cent of compensation paid under Workers Compensation). The Woodhouse Report states that

this division of energy, time and money together with the normal processes of competition makes it inevitable that the ratio of expenses to compensation must be high.

Furthermore, it adds that the system in place was expensive ‘not because the system was mismanaged, but because the system makes this inevitable’. This inevitability arose from the absorption for administration and other charges as much as $40 for every $60 paid over to successful claimants (Woodhouse Report, s216).

Danzon (1990) shows that in the first 15 years of the scheme, overhead costs fell from 50 per cent of total expenditures under the tort system to under 10 per cent using ACC (Danzon, 1990:4). Operational costs in 2001 represented 10.05 per cent of expenditure as classified in the annual report. However, these costs are a significant under-representation relative to the previous insurance and tort-based scheme, as employer and employee levy collection costs under ACC are borne by the Inland Revenue Department, motor vehicle levies are collected by the Land Transport Safety Authority, employer safety monitoring and education costs are borne by the Department of Labour ($25.4 million in 2001), and road safety education spending is the responsibility of the Land Transport Safety Authority ($15.323 million in 2001).

Both public and private health sector providers underwrite safety and education programmes with considerable monitoring components (for example, the Royal New Zealand Plunket Society operates New Zealand’s largest child car seat rental and safety programme, involving education, monitoring and compliance checks, using its own privately-sourced funding). Further, the compliance costs incurred by medical practitioners and employers in administering the scheme are also opaque. Although scheduled reimbursements to medical practitioners are intended to compensate for these costs, the schedule-
based nature of payments (insensitive to the time and expertise required to process the claim) potentially disguise the real transaction costs.

The costs of lost incentives

While ACC may have met its compensation certainty and many of its transaction cost objectives, evidence on costs arising from the loss of incentives is less clear. While counterfactuals are difficult to assemble, there is sufficient evidence for concern that the New Zealand scheme has severely distorted incentives for individuals to take care due to the very limited ability to either punish via the torts system and to the inability to relate levies to risk-based behaviour.

Kessler (1999) clearly identifies the need to increase monitoring to compensate for the removal of loss-preventing tort incentives when no-fault schemes are introduced. In the New Zealand case, in relation to employment injuries, the removal of the right to sue has reduced the incentives for both employers and employees to take precautions. However, the extension of the scheme to encompass all citizens means that the removal of right to sue should have been associated with increased monitoring and enforcement for all accident-related activities for all citizens. This includes additional monitoring and enforcement for all potential employment, self-employment, vehicle, medical misadventure and non-earner related accidents.

Generally speaking, employees can act more riskily in the certainty that they will be compensated for the additional accidents this causes, and employers have minimal incentives to invest in safety equipment and procedures as any additional accident costs are met from the fund and not from individual pockets (Mears and Chapple, 1996:35). The current arrangements include provisions to manage employer incentives through co-payments (the first week’s earnings compensation for injured workers) as well as to enable linkage of employer levies to individual safety records via rebates and penalties (Mears and Chapple, 1996:61), but the latter are extremely limited.

However, incentives on employees are negligible. Consequently, spending on monitoring and enforcement of workplace safety standards should be higher to ameliorate for the loss of the tort right. Yet total occupational health and safety (OSH) spending was only $25.4 million in 2001, equating to slightly over $15 per employee and $97 per business. Thus, for workplace safety at least, there appear to be both negligible incentives and negligible monitoring, enforcement and education in the New Zealand environment. It is not surprising that New Zealand’s workplace accident rates are high on a world basis (Wren, 1999) and particularly high relative to Australia (Department of Labour, 2001:11).

While there is a minimal amount spent via OSH on workplace safety monitoring and enforcement, for the balance of potential claimants on ACC (for example non-workers, sports players, and motor vehicle drivers) neither incentives nor dedicated accident prevention monitoring and enforcement exist outside of bundled offerings provided by Police and the Land Transport Safety Authority dedicated to road safety — $9.583 million in 2001 on policing and $15,323
million on education and safety promotion (LTSA Annual Report, 2001)). While ACC does have an education role, these expenses are indistinguishable from other expenditures in the Annual Report. As they are neither compensation nor rehabilitation expenses, we presume they form part of operating expenses and are therefore substantially less than the 10.05 per cent of expenditure assigned to this function. While it is impossible to determine the actual extent of monitoring, enforcement and education undertaken, adding OSH and LTSA expenditure implies that, overtly, only $50.3 million of Government expenditure (2.5 per cent of ACC expenditure in 2001) is being spent on strategies to directly redress for the loss of the tort right.

**The costs of insufficient information**

Moreover, as Danzon (1990:36) cautions, all monitoring, enforcement and loss-prevention activities require significant amounts of information to be efficiently undertaken. Mechanisms within ACC do not support the collection of information to further develop these roles. Private insurance schemes, however, do have an incentive to maintain information that will enable them to develop and deliver accident prevention programs, and thus provide more efficient compensation schemes. It is debatable whether such information can be efficiently collected, maintained and utilised in diversified, multi-agency environments. This is a very real danger in government-mandated monopoly schemes, over and above the issues associated with the removal of incentives in no-fault systems, and is a contributing factor in rising accident rates.

Potential tort liability also provides a strong incentive to maintain information and use it efficiently to lower accident risk. To quote Danzon (1990:36-37):

> The low overhead rate of the system is achieved in part by foregoing the collection of the data necessary to understand the causes of accidents and implement appropriate prevention plans. … (T)he evidence that private insurers operating in competitive insurance markets elect to keep such data suggests that it is a cost-effective investment. … The state monopoly insurer appears to assign disproportionate weight to minimising overhead, which is the more visible component of the cost, to the relative neglect of the visible but equally real social costs of injuries that could have been avoided had there been better information-based risk management strategies in the first place.

**ACC 1999-2000**

Partially in response to concerns such as those of Danzon (1990) and Mears and Chapple (1996) about the costs of prevention, the worker compensation portion of ACC was opened up to private providers in 1999. These providers could charge employer premiums for worker cover on a fully risk-rated basis and compensate
employees according to the same (or better) terms as those available under the
government ACC schedule. While reintroduction of the right to sue was not
allowed, the reforms partially addressed the perceived low levels of prevention
awareness and action in workplaces. Following the election of the Coalition
Government in late 1999, the legislation enabling this variation was repealed and
the six insurance companies (five private and one government-owned) established
to meet this market were disbanded in 2000.

While the period of operation of the revised scheme was too short to enable
any accurate assessment of outcomes in terms of either scheme cost or accident
levels\(^1\), there was considerable support for the new companies from many sectors.
Anecdotal evidence supports the contention that for those firms opting to insure
with the new companies, heightened priority was given to workplace accident
prevention in the process of negotiating premiums.

**Minimal Incentive, Negligible Monitoring: Who Bears the Risks?**

Given that the New Zealand ACC system appears to have minimal incentives for
the prevention of injury, negligible spending on monitoring, enforcement and risk
prevention, and an apparent absence of political will for any mechanism for
managing moral hazard; the question remains: who bears the costs and risks of
moral hazard actions?

In the first instance, taxation revenue will meet the costs of any shortfall in
revenue between levies and claims. However, if resources for reducing accident
rates are minimal, the net result is that more accidents than is efficient will occur
Kessell (1999) — and this is quite likely to be the case in New Zealand (Wren,
1999; Ministry of Labour, 2001). Total compensation costs will be higher
(necessitating higher calls than necessary on taxation revenues and levies shown
by Mears and Chapple to be, in effect, a payroll tax borne by workers) and greater
suffering will be incurred (as more people than is efficient will endure accidents
and their consequences).

Ironically, when the balance between incentives and monitoring and
enforcement is inefficient, it is to the very people that the scheme was designed to
protect — potential accident victims — to whom the extra risks and costs are
shifted. Those with significant power to alter the probabilities of the accident
occurring (for example, employers via their safety equipment purchases, or
politicians who set the spending limits for monitoring and enforcement bodies)
bear practically none of this additional risk (Howell, 2001). All New Zealanders
are exposed, and few face direct incentives, resulting in a ‘tragedy of the
commons’ — and overall standards of care in all accident-related activities are
inefficiently low.

---

\(^1\) Crichton et al. (2002) are building a matched employer/employee data set in order to
investigate changes in firm and worker behaviours over this period
Lessons from the New Zealand Experience

The New Zealand experience with ACC highlights that the introduction of any no-fault scheme that waives the right to sue cannot ignore the need to increase overt monitoring and enforcement concomitant with the waiver of legal rights, if the optimal level of loss occurrence is to be achieved. While transaction cost reduction and compensation certainty may be an attractive lure to adopt such schemes, it cannot be ignored that efficient schemes must compensate for the loss of prevention incentives with other mechanisms. It also cannot be ignored that these mechanisms require both cash expenditure and collection of the appropriate information to enable them to perform as effectively as the same (or desirable) level of prevention possible via incentives.

The New Zealand experience with no-fault accident compensation, in the absence of tort action to modify moral hazard behaviour, is almost unique. While ensuring certainty of payment, it is far from clear that the scheme has succeeded in balancing the transaction costs and benefits of overt monitoring and enforcement against the costs and benefits of incentive management available from tort action. Indeed, there are indications that the net costs, in lost information and higher rates of residual, uncompensatable injuries, may be high. Whether the brief experiment with individual employer risk-rating during 1999-2000 succeeded in reducing risks and costs is hard to determine. Without a counterfactual that represents current injury-related behaviours in a tortious environment, it cannot answer the questions surrounding the extent of incentive loss due to the absence of the right to sue in all non-employment related accidents.

The lessons from the New Zealand ‘experiment’ appear to indicate that there are very real dangers in removing the right to sue along with no-fault insurance. Indeed, the two regimes together may provide the best of both worlds — certainty of compensation and a cost-effective way of managing moral hazard behaviour. While the right to sue in New Zealand has been removed in respect of personal injury, a form of no-fault insurance has existed in motor vehicle insurance in the agreement between insurers to meet the claims of their own clients and not pursue redress from each other for the actions of the others’ clients. Only where clear fault is established, is redress sought from the individual who caused the loss, and that is of a punitive nature (for example loss of the liable party’s no-claim bonus) rather than recovery of the loss. This potential cost acts as an additional incentive for third parties (akin to employers) to act with due caution, is self-reinforcing, and does not appear to have had any negative effect on either the rate of claims or the costs of accidents. Indeed, it is ironic that greater incentives should be available to prevent loss to property, which can be fully compensated financially, than for persons, where injury may not be fully compensated by monetary exchange.

With respect to increased interest in no-fault schemes, it would appear that there may be a role for such schemes where the ability to manipulate the probability of an adverse event occurring is negligible. However, where the transaction costs of monitoring and enforcement are high, or potential moral
hazard actions numerous and hard to monitor (as in New Zealand’s far-reaching ‘one compensation scheme suits all accidents’ environment), then no-fault schemes probably require additional incentive-based mechanisms, such as tort law, in addition to criminal sanctions. The distinction between an ability to manipulate the extent of loss and an ability to manipulate the probability of the adverse event is crucial to understanding the potential costs of implementing no-fault insurance schemes. Governments contemplating such schemes in the rush to fill a missing market, as a result of recent changes in the types and forms of risks to which we have become exposed, should be aware of the efficiency implications.

References


The authors acknowledge helpful comments provided by Graeme Wells and two anonymous referees in the preparation of this article.
Second and Third Thoughts on Privatisation in Indonesia

Ross H. McLeod

Indonesia’s economic policies began to become much more market oriented during the 1980s. Various policy reforms were implemented, notably in the field of international trade (Fane and Condon, 1996). In addition, there came to be a new emphasis on privatisation, although this was nearly all talk and no action (Hill, 2000:103-5). In 1989 the then Finance Minister announced that 52 state-owned enterprises (SOEs) would be listed on the Jakarta Stock Exchange between 1990 and 1992 (Habir, 1990:101); in the event, almost none were. In 1993, the then Minister for Research and Technology, B. J. Habibie, claimed that a similar number could be sold quickly (McLeod, 1993:7); again, almost nothing came of this. Nevertheless, although there was a conspicuous lack of progress with privatisation as normally conceived, there are several examples of effective privatisation, provided this term is interpreted sufficiently broadly.

Approaches to Privatisation

In the conventional sense, ‘privatisation’ refers to the sale of SOEs to the private sector. The Indonesian government owns a wide range of enterprises, including electricity, water supply, transportation, telecommunications and construction companies, mining and manufacturing firms, plantations, banks, insurers, importers, retailers and hotels. There is, therefore, much scope for privatisation by way of divestment of SOEs.

In a broader sense, privatisation can be thought of as encouraging relatively greater private sector involvement in parts of the economy relative to the public sector. By definition, the divestment of SOEs results in a larger share of the market in question being supplied by the private sector. But this can also be achieved by encouraging the involvement of, and competition from, private enterprises in sectors where state-owned firms exist. There was in fact quite considerable privatisation in Indonesia in this broader sense in the late 1980s and the 1990s. We shall consider examples from three different industries in this paper: banking, civil aviation and electricity.

The Banking Sector

Commencing in 1982 the government began to deregulate the banking sector, which was at that time heavily dominated by seven large state institutions: five commercial banks, a savings bank and a development bank (McLeod, 1999).

Ross McLeod is a Fellow in the Indonesia Project at The Australian National University.
There were many more privately owned banks, but in aggregate they accounted for only seven per cent of total bank assets. In the early phases of deregulation the government abolished its practice of trying to impose ceilings on the growth of each bank’s assets, which had held back the growth of the private banks. In the later phase, commencing in 1988, the government abolished barriers to entry to banking that for many years had prevented the establishment of new banks. In addition, it removed bureaucratic obstacles to the expansion of the banks’ branch networks.

The impact was dramatic (Table 1). The number of private banks increased rapidly, and the number of bank branches even more so. Other indicators, such as the number of savings bank accounts and the total amount of deposits of all kinds grew extremely rapidly as banking became much more accessible to the general population, and as banks offered attractive interest rates and other incentives to their customers. Previously many businesses, and most individuals, had no bank deposits at all. Although the existing state banks also expanded their branch networks and became somewhat more customer oriented, the net result was that the market share of the private banks grew dramatically at the expense of the state banks. In other words, the industry as a whole moved in the general direction of privatisation, even though no state banks were sold. In a single instance, a minority share (25 per cent) of Bank BNI, one of the state banks, was sold to the general public (in 1996). The bank was publicly listed, but still continued to be thought of, and to behave, as a state-owned bank.

### Table 1: Impact of Deregulation on Indonesia’s Private Banks

<table>
<thead>
<tr>
<th></th>
<th>Dec 1988</th>
<th>Jun 1997</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>63</td>
<td>160</td>
</tr>
<tr>
<td>Branches</td>
<td>574</td>
<td>4,267</td>
</tr>
<tr>
<td>Savings accounts (million)</td>
<td>2.6</td>
<td>17.4</td>
</tr>
<tr>
<td>Deposits (Rp trillion)</td>
<td>11</td>
<td>183</td>
</tr>
<tr>
<td>Share of total assets (%)</td>
<td>24</td>
<td>54</td>
</tr>
<tr>
<td>(State banks’ share of total assets, %)(^a)</td>
<td>71</td>
<td>35</td>
</tr>
</tbody>
</table>

Note a: The remaining market share was held mainly by local branches or subsidiary joint ventures of foreign banks.

### Civil Aviation

The case of the civil aviation industry is analogous to that of the banking sector. For decades civil aviation was dominated by the state airline, Garuda. A number of small airlines competed on a few domestic routes, but Garuda enjoyed a near monopoly position by virtue of being the only airline permitted to operate jet
aircraft. This policy was dropped in the early 1990s, at the same time that a controlling private sector interest was established in a small airline, Sempati Air, previously owned by the military. The new policy allowed Sempati to compete effectively with the poorly managed Garuda, just as banking deregulation allowed private banks to compete effectively with the state banks. This new airline quickly established a high profile, and injected a hitherto unknown emphasis on concern for the travelling public. It introduced lottery prizes for passengers, compensated them if departures were late, simplified and speeded up the process of making reservations, and so on. The private sector market share increased rapidly as new domestic routes were added, eventually to be followed by the introduction of a few international routes. The further benefit was to put pressure on Garuda to improve the quality of its own service.

The Electricity Sector
The sale of electricity has always been monopolised by the state electricity company, PLN (Perusahaan Listrik Negara). It should be noted, however, that many of the larger consumers of electricity have their own generation capacity, because PLN has not always had a good reputation for reliability, and it has often been thought more economic to be self-reliant than to run the risk of incurring the cost of power blackouts (McCawley, 1970). In 1990, however, the government announced its intention to encourage private sector involvement in the power generation component of this sector. It contracted with a number of private consortia on a build, own and operate basis to construct some 26 new plants to supply electricity to the national grid. As a result of this, perhaps a third of national power generating capacity (excluding generation for own use) is now in the hands of the private sector. Thus, although consumers of electricity still can deal only with PLN as supplier, to a significant extent the ultimate source is now the private sector.

The Crisis of 1997-98 and Consequent Problems with Privatisation
Superficially, at least, Indonesia seemed to be privatising various parts of the economy very successfully, even if the government made very little progress by way of enterprise divestment. But the process went horribly wrong when the economy went into crisis, along with several other Asian countries, beginning in 1997 (McLeod, 1998).

The Asian crisis brought an abrupt end to the rapid growth of the private sector’s market share in banking. The unexpected float and large depreciation of the Thai baht in July 1997 resulted in a loss of confidence in the rupiah, which lost 24 per cent of its value over the next two months alone. This had drastic implications for the corporate sector, which had faced an exchange rate risk by virtue of having borrowed heavily in dollars rather than rupiah, often in order to invest in sectors producing non-tradables. In turn, many of these dollar loans had been provided by domestic banks, which quickly came under pressure in the form
of large scale withdrawals of deposits. Before long the central bank began to operate as lender of last resort, providing emergency liquidity to banks whose funds were being drained (Enoch et al., 2001:32). This was not enough to stop the run on deposits, however, and the government eventually issued a guarantee of the banks’ liabilities.

Last resort lending is not the appropriate policy when banks are clearly insolvent, rather than merely illiquid, of course, and all that was achieved by it was to delay the inevitable — and to make the eventual losses even greater. Fane and McLeod (forthcoming) estimated the fiscal cost of the claims against the government’s guarantee of bank liabilities to be of the order of 40 per cent of GDP; recent developments suggest that this is likely to be an underestimate.

Sempati Air also became bankrupt during the crisis, with debts vastly in excess of its assets. An important explanation for this is that, since most of its sales were in the domestic market, whereas three of its major cost items — aircraft lease payments, debt service and aviation fuel — were dollar-denominated, it suffered greatly as a result of depreciation of the rupiah, just like many other firms producing non-tradables. Unable to increase its airfares sufficiently — not least because government pressure on its major competitor, Garuda, prevented the latter from raising its fares as well — Sempati quickly found itself facing large negative cash flows that soon caused it to collapse (while government backing of Garuda allowed it to avoid a similar fate). Among its very large debts were unpaid bills for the supply of aviation fuel by Pertamina, the government owned petroleum monopoly, and for the rental of airport facilities of the state-owned company, Angkasa Pura. As a group, the now defunct Sempati Air plus two of its principal owners, the Humpuss and Nusamba conglomerates, is among the largest defaulters on loans from the state banks.

The experiment with privatisation in the electricity industry has been similarly disastrous, resulting in huge losses to the government — albeit much smaller than those from the banking system. In this case the losses are a direct consequence of the nature of the contracts that governed the supply of power by private generators to PLN. These contracts had two crucially important features.

First, the supply price for electricity was denominated in US dollars. This meant that by 1998, after the rupiah price of US dollars had risen by a factor of four or five, the rupiah price of power supplied to PLN had risen in direct proportion. But the political realities were such that the government felt unable to allow PLN to increase its prices to consumers (denominated in rupiah) commensurately. Thus PLN immediately began to lose on every unit of electricity it purchased from the private power companies.

Second, the contracts required PLN to ‘take or pay’. That is, once the private companies had installed a certain amount of productive capacity, PLN was obliged to purchase all of their output, or to pay an equivalent amount if it purchased smaller quantities. In the context of the unfolding economic crisis, many large factories and other users of electricity were cutting back their own output, and so their demand for electricity fell accordingly. This further reduced PLN’s revenues relative to the amounts it was obliged to pay to the private producers. The overall
result has been an enormous drain on the budget, to which PLN’s profits had been a significant contributor in better times.

Ambivalent Ongoing Support for Privatisation

In short, although previously these three examples of privatisation had seemed quite successful, with the advent of the crisis they ultimately proved to be very costly failures. Nevertheless, the government is still committed to ongoing privatisation, especially as this is one feature of economic policy demanded by the IMF in return for its continued crisis-related financial support.

It is worth noting in passing that there has been some ambivalence on this point, especially in relation to the banking system. When it became clear that many of the banks — and all of the major banks — were insolvent, the government decided, in consultation with the IMF and the World Bank, to:

- take over, or at least take a majority shareholding in, many of the large private banks that had failed, and recapitalise them
- transfer the deposit liabilities of banks that were closed to state banks
- set up what was in effect a new state-owned holding company, the Indonesian Bank Restructuring Agency (IBRA), to which were transferred all of the assets of banks that had been closed, along with the worst of all the non-performing loans of the state banks and the private banks that had been recapitalised.

The intention, nevertheless, was for these arrangements to be temporary. Nationalised banks were to be privatised as soon as possible; newly acquired government stakes in private banks were to be divested as soon as possible; the state banks themselves were soon to be privatised; and IBRA was to have a limited life of only five years during which it would sell off all of the assets it had acquired and then be wound up.

There has been a disappointing lack of progress, however. All the state banks are still owned by the government, and their relative importance has grown significantly by virtue of the large volume of deposits transferred to them from failed private banks. Only one of the nationalised banks has been divested.

This lack of progress is partly attributable to policy makers having second thoughts on privatisation: given that the ‘privatisation’ that had been achieved prior to the crisis has turned out to be a failure, it now appears to many that privatisation is not so desirable after all. To argue against privatisation for this reason, however, is conveniently to ignore the fact that state enterprises have fared no better than private during the crisis, and that their survival is attributable to the perceived willingness of the government to repay their debts regardless of whether they are insolvent. For example, the state banks have also amassed enormous

---

1 A detailed record of the evolution of policy during this period is contained in (Enoch et al., 2001:26–40).
losses\(^2\) — as they had done on many occasions prior to the present crisis (McLeod, 1999:281) — as has the state-owned airline, Garuda. Moreover, the state enterprises in general have always generated little by way of profit for the government — or have required large subsidies to keep them going (Hill, 2000:105-7).

Third Thoughts: The True Rationale for Privatisation

Privatisation in the conventional sense is undeniably important in present circumstances from the point of view of providing the cash flows needed to maintain the government’s spending programmes. The government now faces the huge burden of servicing bonds it had to inject into the banks to make good their losses; at the same time, additional borrowing is not an option, and the scope for increasing tax revenues in the short term is very limited. But cash generation is not the real rationale for privatisation. After all, the only sound reason for converting earning assets to cash is if their return is lower than the opportunity cost of continuing to hold them. In the present context, a lower bound for the opportunity cost is indicated by the interest rates paid on bank recapitalisation bonds, which could be retired using the proceeds from privatisation. The true opportunity cost of funds tied up in state ownership of enterprises is somewhat higher than this, since the rates on these bonds are artificially low, having been set unilaterally by the government (McLeod, 2000:28). If private sector owners are more efficient managers, the sale value of the enterprise in question will be greater than the present value of future returns likely to be earned with the government as manager.

Thus the fundamental economic argument in favour of privatisation still stands. The core of this argument is that enterprises will be managed better when managers have their own funds at risk, or are appointed by shareholders with a significant stake in the company in question. The incentives for sound management of state-owned enterprises in Indonesia are largely non-existent: the rule of law is very weak, and accountability of the government through the political processes — given an unsophisticated electorate unused to democracy — is also lacking.

Unfortunately, however, this argument sits uncomfortably beside the reality of the failed experiments with privatisation described above. Evidently it is necessary to think more carefully about the process, and to ensure that the way in which it is approached is consistent with this core argument. We need to analyse Indonesia’s experiments with privatisation more carefully. It is not enough to observe that they have failed; it is necessary to explain why they did so.

\(^2\) Indeed, Fane and McLeod (forthcoming) argue that the state banks’ losses per unit of pre-crisis customers’ deposits were considerably higher than those of the private banks.
Analysing Indonesia’s Approach to Privatisation

Banking: partial divestment of a state bank

It was mentioned briefly above that, of the seven state banks existing prior to the crisis, the government had only divested a small minority shareholding in one of them, Bank BNI. Little, if anything, was achieved by this. The government was not seriously in need of additional cash at that time, and no single private shareholder acquired a large enough stake to be able to exert any significant influence on the policies and strategies of the bank. Presumably the government obtained a price per share from the divestment lower than its potential, since buyers would have realised that the introduction of new management capable of extracting greater value from the bank’s assets was not in prospect. Bank BNI gained some notoriety during the crisis by virtue of having provided well over US$1 billion of loans to a single conglomerate, Texmaco, which is now by far the single largest debtor to IBRA by virtue of defaulting on these loans (Fane, 2000a:29-30). It was revealed in 2000 by the then Minister for State Enterprises that these loans had been provided by instruction of the former president, Soeharto. In other words, having a significant but fragmented minority private sector shareholding in the bank did nothing to offset the debilitating influence of the politically powerful.

It is clear what went wrong in this example of ‘privatisation’. Partial divestment of the government’s shares did not result in any private owner having a sufficiently large stake in the bank to provide the incentive and the ability to exercise strong influence on its management. The managers of the bank continued to have weak incentives to strive for high profits relative to their incentives to act in the interests of other parties such as political patrons, favoured borrowers, friends and relatives.

Banking: private sector expansion

In the case of privatisation of the banking sector by allowing the private sector to expand its market share at the expense of state banks, the explanation for the unsatisfactory outcome of the experiment is also closely related to the failure to ensure that the approach was consistent with the core argument in favour of privatisation. Specifically, the government failed to ensure that the owners of the private banks had significant amounts of their own wealth at risk. There were two reasons for this.

First, the prudential regulations called for banks to observe a minimum capital adequacy ratio (CAR, the ratio of capital to risk weighted assets) of only 8-9 per cent. This reflects an extremely high gearing ratio (that is, ratio of debt to equity) relative to the norms of the corporate sector outside banking. The only

---

3 The minimum CAR was in the process of being raised in stages from eight per cent to 12 per cent. It became nine per cent in September 1997, just as the crisis was getting under way.
obvious explanation for creditors of banks — primarily, their depositors — being willing to accept such a small level of equity to protect themselves against potential losses incurred by the banks is that there is a perception that the government will step in to provide such protection if necessary. When the crisis struck this is precisely what it did, despite frequent denials previously that it provided any kind of guarantee of the safety of bank deposits.

As we have seen, the banks made skilful use of this perception of an implicit government guarantee by actively marketing their deposit products, opening many new branches and offering high interest rates and other benefits such as lottery prizes. The implicit government guarantee amounted to a subsidy, and the more deposits that could be mobilised, the greater its total value. In turn, much of the subsidy found its way to borrowers, who were able to borrow at lower rates than if it had not existed.

Second, the government permitted banks to lend to affiliated parties — roughly speaking, to companies owned by the same people. The prudential regulations attempted to limit this practice by imposing a ceiling on loans to affiliates as a proportion of banks’ capital. The constraint was lax in itself by international standards, but in any case, banks flouted these lending limits with impunity: the central bank, in its capacity as prudential regulator, seemingly did nothing to enforce its own regulation.

The effect of banks lending to their own affiliates is to reduce the amount of funds of the owner genuinely at risk. Suppose that an owner subscribes $1 million as equity in the bank, but then makes a loan to himself of $1 million. If he fails to repay the loan there is then effectively no equity cushion for the benefit of the bank’s creditors if it gets into trouble. If the bank were to be wound up, the depositors would need to try to recover the loan from the owner through the courts, but this may be exceedingly difficult — especially if the loan documentation is inadequate and collateral of low value or non-existent. In a country in which the legal system is inefficient, lacking in competence, and highly corrupt, the chance of success would be negligible. The owner could expect to escape with his million dollars intact and, given the extreme political pressure on the government to protect depositors, the effect would be to shift the banks’ losses to taxpayers. This precisely describes what has happened in countless cases in Indonesia during the crisis.

Civil Aviation

Disparate though the two sectors may be, the failure of privatisation in civil aviation occurred for much the same reason as that in banking — namely, the failure to ensure that the private sector owners of Sempati had a significant amount of their own wealth at risk. In this case there also appears to have been what amounted to an implicit government guarantee that allowed the company to build up large and practically unsecured debts to its major suppliers and
Alternatively, it could be said that the losses to the state in this case derived from the poor business practices of various state-owned enterprises: the state banks that lent to Sempati; the state petroleum company that supplied it with fuel on credit; and the state-owned airports that failed to collect rent payments from it. In this sense, the failure of civil aviation privatisation stems not from the new private sector competitor, but from grossly deficient management (no doubt the result of political interference) of state enterprises with which it had significant business relationships.

**Electricity sector**

Although the task of power generation was partly turned over to the private sector, again this was not done in a manner consistent with the core argument in favour of privatisation. The private sector members of the various consortia that constructed power generation plants did have large amounts of their own funds at risk, but two of the major risks themselves were shifted to the government — consistent with the notion of ‘privatisation of profits and socialisation of losses’. First, there was the risk that the demand for power from each new plant would be significantly less than its supply capacity, either because aggregate demand for electricity had been overestimated, or because of the large number of new generators being constructed. By accepting ‘take or pay’ contracts in its negotiations with the new private operators, the government allowed this risk to be shifted to itself. Second, there was the exchange rate risk, reflecting the possibility that the government might not always persevere with its policy of slowly but steadily depreciating the rupiah against the US dollar (McLeod, 1997:33). By virtue of PLN’s acceptance of dollar-denominated prices for electricity, this risk was also shifted to the government; the outcome of negotiations over contract details no doubt had much to do with the identity of the domestic partners in each of the consortia, all of whom were either relatives or cronies of the president (Thirwell, 2001). In short, the requirement that the risks of the business in question would be carried by its private sector owners was violated in these two very important respects.

One of the principal reasons why government-owned businesses make losses is that they are always under political pressure to hold their selling prices down, even though their costs may be rising. There are also pressures to hold buying prices (such as wages) up, and to extend excessive amounts of credit for lengthy terms and without adequate security. In the private sector, firms are generally free to put up their prices and, although this may meet with opposition from their customers, there is little to fear except if the price increases are intended to offset inefficiency or to earn excessive profits. In these cases the firms can expect to lose market share to more efficient and less avaricious competitors, but private sector firms are far less vulnerable than SOEs to interference from politicians.

---

4 In Indonesia’s weak and corrupt legal environment it has been common for borrowers to divert cash revenues to related companies in such a manner as to be unrecoverable by creditors. It is noteworthy that Sempati was declared bankrupt at its own request, not that of its creditors.
worried about being voted out of office if prices increase. Thus privatisation of power generation in Indonesia can be said to have failed because the approach followed did nothing to deal with this intrusion of politics into commercial decision-making, resulting in the transfer of financial risks from the private producers to the government through PLN.

Making Privatisation Work

Banking

The government’s stated intention is to return ownership of the wholly or largely nationalised former private banks to the private sector, and to sell off the state banks. Obviously it is important that the previous mistakes with privatisation are not repeated. Action is required on several fronts to avoid the kinds of outcomes described above, by ensuring that private bank owners have significant amounts of their own capital at risk.

First, much higher standards for capital adequacy need to be set. It is arguable that the CAR of eight per cent prescribed in the international convention known as the Basle Accord (Basle Committee on Banking Supervision, 1988) is too low, even in developed economies. An important purpose of capital, or equity, in all businesses is to provide a cushion of safety to creditors. Clearly this cushion needs to be larger when the risks are relatively large, which is certainly the case in Indonesia (and probably in developing countries in general), where political stability cannot be taken for granted, the legal system is exceedingly weak, and the economy is vulnerable to negative shocks of various kinds. On this basis, Fane (2000b:127-30) argues in favour of a CAR of at least 16 per cent for economies similar to Indonesia.

Second, much greater care needs to be taken in measuring and defining capital for the purpose of calculating the CAR. A more realistic approach is called for, in which assets are marked to market frequently, and adequate provisions for loan losses are created in timely fashion whenever there are indications that particular loans may not be repaid in full. In addition, capital should be defined to exclude the value of loans or any other kind of exposure of the bank to affiliated companies. In the previous example of a bank with equity of $1 million that lent $1 million to a firm controlled by its owner, the bank’s capital would be recorded as zero for regulatory purposes, because the $1 million loan to its owner would be offset against the $1 million of equity. This would have the desirable impact of removing the incentive to set up banks for the main purpose of obtaining cheap finance for affiliated companies by exploiting the implicit government guarantee of bank safety.

Third, the prudential regulator should be far more active in requiring corrective action from banks whose capital adequacy falls below the specified minimum. There should be very little regulatory forbearance: owners should be required to repair any capital deficiency without delay. If they failed to do so the bank should be taken out of their control and sold to another party, who would be
required either to build up the bank’s capital immediately or to liquidate it. If the bank continued to operate, the original owners should receive the purchase price paid by the new owner. If it were liquidated, they should receive nothing (Fane and McLeod, 2002).

Fourth, given the apparent impossibility of persuading the general public that the government will not make good their deposits in cases where banks fail, it would also seem sensible to implement a system of deposit insurance, provided it is operated on genuine market principles. Banks would need to pay insurance premia that reflected the insuring institution’s assessment of the risk of failure of each bank: the insurer should not be subsidised in any way by the government. Under such arrangements it could be expected that insurance premia would be inversely related to capital adequacy: the higher the bank’s capital relative to risk assets, the lower the risk to the insurer, and the lower the insurance premium. This would provide a market-based incentive to maintain higher CARs, rather than relying purely on enforcement by the prudential regulator. Indeed, with deposit insurance it would be sensible for the insurer to become, in effect, the prudential regulator, since it would have a relatively strong incentive to enforce prudential regulations: these would simply become part of the terms and conditions of the insurance contract.

There is no reason why the deposit insurance institution or institutions should be owned by the government. Indeed, given Indonesia’s long history of poor management of state-owned financial institutions (not least as a result of heavy political interference) there is a strong argument for requiring them to be owned by international entities of strong repute, or simply to require the banks to purchase insurance from such entities abroad (Schmulow, forthcoming). In this manner, the government could avoid almost the entire need to be involved with the prudential regulation of banks, and it could almost entirely remove the risk of having to bail out banks’ depositors again: it could simply require each bank to be properly insured by an approved foreign insurer (and draw up a list of approved institutions for that purpose). This would be a genuinely market-oriented approach to privatisation of the banking system, similar to the recent emergence of insured investment funds in the Australian capital market.

**Civil aviation**

The first best policy here would be for the government to fully divest itself of Garuda and to open up the industry to new competitors, including foreign airlines; there is no sound justification for having a government-owned airline. If the government thinks it necessary to have aircraft serving particular routes that would not be commercially viable it can easily provide an explicit subsidy to a private airline to do so — and justify this policy to the electorate. In addition, it will be essential to ensure that other government entities provide no hidden subsidies, and avoid significant financial exposures, to private airlines. For example, Pertamina should insist on some arrangement such as regular weekly payments for fuel not more than, say, one month in arrears; likewise the airport management company
should insist on payment of rent in advance for premises and other facilities on, say, a monthly basis. (Better still, Pertamina and Angkasa Pura should also be divested. And if the plan to fully privatise the banking system comes to pass, then of course there will be no possibility of loan defaults by airlines at government banks.)

Electricity

A range of issues relating to future private sector involvement in the electricity sector need to be addressed if this is to be successful. The most immediate question is whether there is really an actual or incipient shortage of power generation capacity, as has been suggested often in recent times (McBeth, 2001). This is such a marked reversal from the story early in the crisis — when the major concern was with excess generating capacity after dramatic expansion of private facilities — that it seems more likely a reflection of artificially boosted demand resulting from a significant reduction in the real price of power to consumers (because of the government’s unwillingness to raise prices in line with cost increases). If this is the case, the appropriate policy change is to adjust prices upward immediately in order to reflect the real cost of power production.

A second desirable change in relation to electricity generation would be for PLN’s own existing production capacity to be sold to the private sector. Just as with banking and civil aviation, there is no sound reason to have a state-owned firm competing with private firms in power generation. In the course of divesting existing generators the government would need to come to grips with the appropriate form of contracts for the purchase of electricity by PLN. It is beyond the scope of this paper to canvass the wide range of possibilities (which also encompass divestment of the national grid and handing over the retailing function to the private sector); suffice it to say that a major objective should be to ensure that the commercial risks of producing power are borne by the suppliers.

One benefit of divesting existing plants (through an open tendering process) is that this would reveal a reasonable supply price, or set of prices, for electricity. This could provide a basis for renegotiation of the contracts with existing private producers. These contracts not only shift major commercial risks to the government, but also specify prices well in excess of those that can be observed in other markets, reflecting the failure of the then government to ensure competitive tendering for the right to become a supplier to PLN and, related to this, the presence of Soeharto cronies and family members in all the joint venture consortia involved (Bosshard, 2000). The present government has already found that attempting to renegotiate these essentially corrupt contracts is fraught with danger, as the foreign partners are sufficiently influential to be able to rely on the support of their own, very powerful, governments in the US, Japan, Germany and the UK (Murphy, 1999). Nevertheless, its political position will be considerably

5 PLN has already talked of moving in these directions (Motoyama and Widagdo, 1999:10-11), though its plans seem naively optimistic given Indonesia’s lack of achievement hitherto with privatisation.
strengthened if it can present strong evidence that the existing contracts are grossly overpriced.

Conclusions

The divestment of state enterprises is not the only way to privatis e industries. In Indonesia, privatisation has involved processes by which private sector involvement increases significantly without very much state enterprise divestment, despite the wide scope for it and despite the stated intention of successive governments to undertake such divestment. Considerable gains can be achieved by encouraging private sector involvement in this manner, whereas achieving similar gains simply by transferring ownership of existing state enterprises to private owners might face significant political obstacles.

Nevertheless, the experience of the last few years has shown that the devil is in the detail so far as privatisation is concerned: there are very significant dangers from privatisation, of whatever kind, if the core rationale for it is not kept clearly in mind. Specifically, the private sector is not necessarily more likely than the public sector to ensure that the enterprise is managed well unless the owners have a significant amount of their own funds at risk, and unless the major risks of the enterprise remain with its owners rather than being transferred back to the government in some manner. In short, Indonesia’s second thoughts on privatisation are well justified. Nevertheless, the correct conclusion to be drawn from recent experience is not that privatisation should be abandoned, but that implementation in a manner consistent with its fundamental rationale is essential.

References


Fane, G. (2000b), *Capital Mobility, Exchange Rates and Economic Crises*, Edward Elgar, Cheltenham, UK and Northampton, USA.


The author is grateful for helpful comments by two anonymous referees.
From an international relations perspective, this book attempts to make the case that the ‘realism’ paradigm which has been dominant in thinking about international security is no longer appropriate, particularly in the East Asian region. Dupont argues that transnational issues such as environmental degradation, unregulated population movements, and organised crime are not adequately captured by the realism paradigm that focuses on the balance of power between states. It is argued that these transnational issues are particularly important in East Asia and the realism paradigm needs to be at least broadened to take them into account.

In the realism paradigm of an unending struggle between states over the protection or acquisition of territory, military threats are paramount and factors affecting military strength are important only to the extent that they affect the military power of the state. For example, there may be concern with competition for resources such as minerals or oil, as these may affect the economic and military power of the state and thereby give rise to conflict. Resource scarcity in an East Asian context is seen by Dupont as having three characteristics that sets it apart from traditional thinking about resource competition: first, it involves resources such as water and air that are fundamental to survival and were once considered to be inexhaustible; second, resource vulnerabilities are being made more acute by the degradation of the natural environment; and third, while the growing scarcity of renewable resources is aggravating tensions between states, for the most part it is the internal cohesion of the region’s states that is most threatened by the prospect of food, water and energy shortfalls, organised crime, unregulated people movement, and AIDS.

Following an introduction which argues the case for the inappropriateness of the ‘balance of power’ paradigm for analysis of security in the East Asian region, Chapter 1 makes the broad case for the importance of the transnational issues to considerations of international security generally and in East Asia in particular. The remainder of the book divides into three parts. Part I examines the importance of rapid population growth, deforestation, pollution, climate change, and reduced access to food, energy and water as drivers of intra-country and regional conflict. Part II looks at recent trends in the unregulated movement of people across borders, including the increasing role of people smugglers, and their likely implications for intra-state and inter-state conflict. The final Part III documents the growth of organised crime and its transnational activities.
particularly drug-trafficking. The HIV/AIDS pandemic is seen as being promoted by the illegal trade in drugs and sex.

Dupont is obviously sympathetic to the ‘deep green’ arguments. He sees the projected global population growth on the one hand and the slowing rate of increase in agricultural (mainly grains) productivity on the other hand as pointing to growing food shortages. This widely accepted conclusion, which has been derived from analysis that focuses on the supply side and ignores factors affecting demand, is wrong. It ignores the fact that population growth is slowing rapidly and a large proportion of the world’s population has experienced increased incomes, and has therefore been through the most rapid growth phase in their food consumption. Therefore, agricultural productivity does not need to increase as fast as it did previously to match effective food demand.¹

Dupont claims that agricultural productivity has slowed because of environmental degradation, particularly the erosion of soil and soil fertility (p. 91). Moreover, it is claimed that soil erosion is three to eight times higher in East Asia than the world average (p. 58). Such claims are highly suspect. Crosson and Anderson (1992) who have carefully reviewed claims about the extent of soil loss and Lindert (2001) would suggest that there is no empirical basis on which to make such a claim.

Similarly, it is incorrect to argue that ‘High levels of population growth can erode the benefits of economic development and worsen income differentials’ (p. 42). Such a statement shows a lack of understanding of the causal relationship between population growth and income growth. Causality runs from income growth to population growth, not in the other direction. If incomes are increasing, the fertility rate and population growth rate decline. Dupont also misunderstands the dynamics of the global demographic transition. The world did not experience the unique event of the rapid population growth of the second half of the 20th century because of a ‘dramatic increase in human fertility rates’ (p. 36), but because infant mortality and adult death rates declined while fertility rates also fell, but with a lag.

Dupont’s lack of understanding of the global food system leads to claims that ‘current world grain production averages around 200 million tonnes’ (p. 96) and that ‘by 2025, China may have to import as much grain as the world produced in 1998’ (p. 83). Dupont is confusing world grain trade, which is about 200 million tonnes, with world grain production. China already produces around 500 million tonnes of grain.

Dupont’s tendency is to argue that the long-term declines in the real prices of food and energy commodities do not give a good indication of real price movements in the future. If he has any confidence in his predictions about demand and productivity increases, he should be investing in these commodities in the expectation of price increases. For example, he is concerned that the rate of crude oil discoveries is continuing to fall when global primary energy use is ‘conservatively expected’ to double or triple by 2050. Clearly, those investing in

¹ See Duncan (1998).
oil exploration and development do not believe such demand forecasts and are expecting energy prices to continue trending downwards in real terms.

In contrast, the author is happy enough to view an increase in fish prices as a ‘sure sign of scarcity’ (p. 102). The increase in international disputes over fishing is also taken as a sign of increased scarcity. There is no thought given to the fact that the Law of the Sea, which gave property rights to countries over the seas within 200 miles of their coastlines, was only ratified in 1994. This allocation of property rights over what were previously ‘open seas’ obviously upset many historical fishing patterns. It will take some time for countries to devise ways to take control over the allocation of these fishing rights and to monitor and enforce them effectively. There also remain disputes over the boundaries of these so-called exclusive economic zones. Fishing grounds can clearly not be treated on an ‘open access’ basis. Hence, the disputes over fishing rights can be seen in a more positive light as the working out of effective controls over what were previously ‘global commons’.

Indeed, many of the transboundary issues that Dupont is concerned about should be seen in the light of cooperative international efforts to solve these problems. Rather than being a growing threat to international security, their identification may be promoting international cooperation to solve them. As well as the Law of the Sea agreement, the Agreement on Cooperation for the Sustainable Development of the Mekong River Basin, the Kyoto agreement, and the international agreements on limits to whaling and fluorocarbons are efforts at an international level to resolve transboundary problems.

Scarcity of fresh water is, I believe, correctly identified as the primary threat to international security. The author emphasises that appropriate pricing of water is necessary to reduce the enormous waste and to allocate water to its most productive use. The sharing of major river and groundwater systems between nations has proven to be a most difficult problem to resolve. As far as management of water flows is concerned, however, the author is clearly in the ‘anti-dam’ camp and does not give any consideration to the incredibly important role that dams have played in the provision of fresh water — the greatest contribution ever to human health. Also, as far as East Asia is concerned, the author notes the great fluctuations in supply of and demand for water, yet does not appreciate the role that dams have in managing these fluctuations.

On unregulated population movement (UPM) Chapters 7 and 8 provide salutary reading for those who believe that Australia is an outlier in its efforts to control its border. People smuggling in order to avoid entry controls is clearly a major international activity and efforts are being made by both rich and poor countries to control it. However, whether unregulated movements of people are likely to be a major cause of international instability is another matter. Dupont acknowledges that the link between UPM and rapid population growth, environmental degradation, high rates of urbanisation and pollution and, in turn, with international conflict is not well established, but is happy enough to accept it. North Korea is used as an example to justify his position (p. 166). But North Korea is such an outlier in terms of international norms of behaviour that it would
Transnational criminal organisations (TCOs) are seen as posing a threat to national security as they are undermining the political authority of governments through challenging the state monopoly over taxation and violence. International security concerns raised include drug and arms trafficking, smuggling of nuclear material, illegal migration, money-laundering, and technology crime (p. 183). Chapter 9 provides an interesting description of criminal organisations around the world. However, there is no theoretical foundation provided for their existence. Nor is there a convincing argument as to why they pose a threat to regional security. It is argued that they pose a threat because they subvert the political authority of the state and imperil its economic and social foundations. But this is a tautological statement. Dupont in fact argues that criminal organisations are tolerated, even encouraged, by governments, and gives the examples of North Korea, Burma, and even Japan. Why is this? Mancur Olson’s (2000) analysis of governments as ‘stationary bandits’ may throw some light. Maybe organised crime is tolerated as part of the taxation system. But Olson argues that ‘stationary bandits’ do not over-tax their subjects. Thus, it might be argued that organised crime will not go past a certain point in its collection of ‘taxes’ so that it will not destroy the economy through over-taxation or lead to the creation of tensions that give rise to international conflict.

There is a chapter devoted to a discussion of the development of the HIV/AIDS pandemic in the region and efforts to control its spread. The spread of AIDS is linked to organised crime and its involvement in trafficking in drugs, sex and people. It is argued that the pandemic could destabilise states but there is no evidence for this yet. As far as national and international security is concerned, the claimed infection rates of military forces of two to five times higher than the general population is a concern. AIDS also has the potential to weaken states economically through the debilitation of its working age population and the high costs of medical treatment.

In assessing the book’s contribution, how justified is the claim that the realism paradigm is no longer appropriate, or that it is not appropriate for East Asia because of the increased importance of these so-called transnational problems? Even the author does not wish to take the argument too far, as he ends up concluding that ‘balance of power’ considerations should still be the ‘core’ of the analysis of national security, and that these transnational issues are ‘outer core’ factors (p. 231).

How important may the various transnational factors be in terms of weakening nation states and leading to conflict? In particular, how valid is the claim that they are ‘reshaping East Asia’s security landscape’ and ‘will play a seminal role in determining East Asia’s future security environment’ (p. 12)? Dupont himself notes that most armed conflicts since 1989 have been internal. Hence, it is difficult to argue that it is the transnational dimension of factors such as population growth, resource scarcity, organised crime, etc., that is the problem.
Even within states, Dupont’s concerns about population growth and environmental degradation leading to increased problems of access to food and energy are seriously in doubt. Certainly, problems such as organised crime, drug use and AIDS can adversely affect economies. But there is no evidence presented of these factors posing problems for national or international security. True transboundary issues, such as air pollution, access to fresh water and fishing, are areas where use has grown to the point that these resources can no longer be treated as though they are in unlimited supply. Hence, their use has to be allocated between countries and between individuals. But instead of these issues being seen primarily as potential sources of international conflict, they are in reality being managed more or less effectively through international cooperation that could be considered to be strengthening international security. Access to water involving two or more countries appears to be the most difficult issue to resolve in an equitable and efficient way. Hopefully, this problem is not beyond the wit of man to solve.

References


*Ron Duncan is Professor and Executive Director of the Australia Pacific School of Economics and Management at the Australian National University.*
Regulation of Electronic Theft


Reviewed by Daniel Stewart

Most definitions of theft involve the dishonest appropriation of property belonging to another with an intention to permanently deprive them of it (Bronitt and McSherry 2001:674). *Electronic Theft* however goes much further in describing the instances in which electronic material and media are involved in forms of unauthorised appropriation. Indeed, much of this book is about recognising new boundaries — in defining ways in which established crimes may be committed, in recognising new activities as criminal and deserving of regulation, and in the development of new conceptions of how that regulation may be achieved. This book attempts to describe the ways in which these boundaries are being shifted, and illustrates the challenges and opportunities facing regulation in this area.

As the book suggests (p. 2), ‘the fundamental principle of criminology is that crime follows opportunity, and opportunities for theft abound in the Digital Age’. Many of these opportunities are not new: electronic media often merely facilitates the carrying out of the forms of acquisition or extends the potential reach or effect of such activities. Thus the ability to steal funds and threaten disclosure of information is increased through the ability to access and manipulate their electronic storage. Similarly, conveying distorted or deceptive information is easier with increased access to new forms of communication and is potentially more profitable due to an increased audience. This book provides many, almost titillating, examples of how the increased reliance on electronic media and storage by private and government bodies has brought with it an increased exposure to fraud, extortion or theft.

Apart from increasing the potential efficiency of criminal activity, electronic media have also contributed to making intangible forms of property an attractive target. Unauthorised, or at least unpaid for, access to computer software, electrical, telephone or internet services has accompanied the development of these facilities, and in some cases required extension of the criminal sanctions available. However the real value of such services is the increased ability to communicate, manipulate and make use of information. The theft discussed in the book largely involves appropriating that ability as much as knowledge of the information itself. The theft of electronic funds often involves access to passwords or other forms of identification or security information, extortion is based on the ability to access information and make it public, deception is premised on ensuring there is inadequate access to or use of accurate information. Protection of the dissemination of information and relationships of confidence through intellectual property has taken on new importance.
Perhaps the most important area exposed through digital media is the privacy of personal information. The ability to collect, collate and take advantage of personal information has been dramatically increased through the large variety of electronic forums in which such information is being gathered. Such information can then be used to permit and encourage theft, provide the basis of extortion or increase the likelihood and value of deception. But it has also given rise to relatively new concerns about the value of anonymity and pseudonymity. Private profiles have become publicly accessible but are still inevitably incomplete and selective, which brings with it the erosion of autonomy and opportunities for discrimination, censorship and unjustified recrimination.

The flip side to any debate about privacy, however, is questions about enforcement, and the ability to identify and locate those who are carrying out regulated activities. Many of the criminal opportunities suggested by the development of digital technology stem from the possibility of at least apparent anonymity. Sophisticated techniques may sometimes track down individuals who break in to computer systems or communicate threats, although the global nature of electronic media means this may often require cooperation between regulatory agencies and countries. But sophisticated techniques are often not practical to prevent less targeted forms of acquisition through deceptive conduct or the misplacing of trust.

Questions of enforcement are just some of the questions involved with the regulation of electronic theft. *Electronic Theft* examines this regulation through the lens of legal pluralism, drawing on an increasing body of scholarship that illustrates regulation as the relationships between various institutions, public and private. Criminal sanctions and government regulation are only two of the influences on behaviour that may be directed towards achieving a regulatory outcome. While traditional forms of public prohibitions and penalties continue to play a role, the government has also recognised the potential effects of providing information and facilitating, perhaps even enforcing, the development of guidelines and codes of conduct. The potential profitability that comes through establishing a reputation and trust in the digital environment has suggested a growing reliance on competitive forces and self-regulation through voluntary codes of conduct and forms of certification. Similarly, many of the features of the new technology that create the opportunities for abuse, such as its accessibility and anonymity, also provide access to encryption and detection technology and other security measures that can help limit vulnerability to electronic crime.

It is in describing how government, non-government and individual actions have all been used to regulate the potential for electronic theft that this book makes its most important contribution. Through pulling together examples from various primary and secondary sources on a wide variety of topics, the authors illustrate the way in which various elements of the conduct being discussed is regulated through one or more of the mechanisms presented.

For example, the chapter on industrial espionage sets out the way in which digital storage, access and communication increases the vulnerability of organisations to breaches of confidence. Outsourcing of technological functions
or specialisation of employment roles increases any exposure. The chapter summarises the way in which relationships of confidence can be protected through the courts, and in some jurisdictions such as the US through criminal sanctions, but emphasises how organisations themselves have the motivation and the means to better establish and protect confidential information. Thus compiling an inventory of confidential information, restricting access, contractual clauses, marking information as confidential and the selection, screening and socialisation of those with access can all assist in maintaining the value of confidential information. But within any organisation this can not be an unqualified objective, and other sometimes-conflicting goals such as efficiency, trust, morale and creativity also have to be sought. The chapter also briefly considers some normative issues about the extent to which confidential information should be protected, given the possible interference with the flow of information available to others and the creation of value through contextualising the information. However, the extent to which these issues can be resolved or even investigated in a principled way given the present interaction of the plurality of regulatory mechanisms is largely beyond the scope of the book.

Clearly the examples presented by the authors are not intended to be comprehensive or detailed. Many more specific forms of analysis are referred to in the text, including some important contributions by one or more of the authors. However, those presented are sufficient to illustrate a number of issues that remain to be explored before any normative analysis of this pluralistic approach to regulation can be undertaken. In discussing legislative responses the authors suggest the need for technologically neutral language and point out the inevitable delay in reactive legislation, but do not explore the pro-active ability of principle-based rather than proscription legislative drafting. The capacity of other forms of judicial if not criminal redress, such as trespass and assault, to adapt to digital environments is also not considered in detail. The difficulties of jurisdictional boundaries are pointed out, but while there is clearly scope for greater cooperation among regulatory agencies, the achievements so far, further impediments and the possible disadvantages of extra-territorial legislation are not always consistently addressed. The relationship between privacy and enforcement, and how boundaries can meaningfully be drawn between them, is not discussed to any great extent.

Any book that draws on the rapidly changing nature of its subject matter for some of its conclusions is itself vulnerable to changes in the regulatory and technological landscape. Since the book was written the Privacy Amendment (Private Sector) Act 2001 (Cth) has been passed, the Australian High Court has seemingly opened the door to a more general right of privacy being developed in Australia while limiting its application to corporations (Australian Broadcasting Corporation v Lenah Game Meats Pty Ltd (2001) HCA 63) and the first charges have been brought in the US for circumventing a technological protection measure. The authors could consistently have used these and many other recent events within their discussion, but the examples also emphasise the unresolved issues inherent this area. The difficulties of establishing limits for privacy
protection and accountability of the means of enforcing those provisions, the uncertain role of the courts in responding to new technologies and protecting relationships of confidence, and what limits should be placed on the ability of individuals and organisations to prevent access to information and expression are issues that are merely raised by the authors.

*Electronic Theft* highlights the inadequacy of trying to examine the issues confronting regulation of electronic media without considering the interaction of the various forms of regulation involved. As the authors suggest, there is much to be done in cataloguing the various forms of institutions, public and private, and the influence that they have. ‘We must now inquire what institutional form, or, even more appropriately, what blend of institutional forms, is best suited to a given task’ (p. 206). Perhaps even more important is the development of new methods to determine the effects, limitations, accountability and equity of those blends. *Electronic Theft* may not push many boundaries of research or analysis, but it clearly illustrates new opportunities for the regulation of electronic technology.

**Reference**


*Daniel Stewart is a Lecturer in the Faculty of Law at the Australian National University.*
The Education of Economists


Reviewed by Colin Rogers

David Colander is well known for his 1991 book titled *Why Aren’t Economists as Important as Garbagemen?* which presented his views on where graduate training in economics in top US graduate Schools was going wrong. In the book under review Colander refines that theme by examining how academic economists came to render themselves less socially relevant. He (pp. 5-6) is concerned that:

In playing the academic game, economists often lost sight of the simple fact that economics was primarily about policy, not logical games. … To me, the correct social goal is for economists to provide as much insight into policy issues, and to understanding the economy, as possible.

His thesis is summarised in the title of the book. Put simply, Colander argues that academic economists have lost the art of economics. Instead they have been seduced by the model of standard science to concentrate on model construction to the almost total exclusion of knowledge of real-world institutions. Consequently, much of what academic economists do makes only a token gesture to policy application. In Colander’s view this is far from adequate as applied policy analysis always requires intricate knowledge of real world institutions. Furthermore, once such institutional knowledge is taken on board it readily becomes apparent that much of what passes for sophisticated theoretical and empirical work either misses the point or involves spurious precision.

To make his case Colander has arranged twelve of his previously published papers into six sections and added an introduction placing the papers in context. Parts I through IV present his vision of the art of economics while parts V and VI ponder the implications of the lost art of economics for the profession and its future. Conveniently, the chapters are generally self-contained and can be read selectively. I will dip into a few to convey the flavour of Colander’s position before I offer an assessment of his thesis. As this journal is concerned with policy I will concentrate on those aspects of the work of direct interest to policy makers. In addition I will outline Colander’s views on the education of economists.

To begin, we had better set out what Colander means by the art of economics. Colander follows John Neville Keynes (1891), and applies a tripartite division to the methodology of economics. By contrast with the familiar positive-normative distinction he introduces a third category called applied economics. The methodology used in applied economics is said to be different from that which applies to both positive and normative economics because it requires knowledge of institutional, operational and social factors. Some of the latter may be non-
quantifiable and acquired only by experience. The methodology of applied economics Colander labels the art of economics and for guidance in the application of that art he sets out six methodological rules in Chapter 3. But rather than examine those rules in detail it is more informative to consider his discussion of the art of monetary policy which originated with an earlier book (Colander and Daane, 1994).

Colander (p. 62) opens his discussion of the art of monetary policy with the bald statement that; ‘Policy works because it has artists pulling the strings’. He goes on to explain that to practice the art of economics it is necessary to understand both the highest levels of theory and the most minute operational detail. ‘The need for both policy makers and academic economists who write about policy to have this dual understanding is the central theme of this book.’ Colander makes a strong case that currently many academic monetary economists fail this test. In particular they lack the institutional and operational knowledge to generate relevant theoretical models. The implication for the education of economists is that many are being taught that these models are more important than operational detail.

To drive home the point Colander examines the academic debate over rules vs discretion in monetary policy and argues that the ‘proof’ of the superiority of rules over discretion does not stand up. He (p. 65) concludes that:

The art of monetary policy is in finding the appropriate rule of limited contingency and in distinguishing those situations in which a rule can usefully be broken because the situation is a sufficiently unique historical event, or one with such long-run consequences that there will be no long run unless the rule is broken —from those situations that can be dealt with by rules.

For Colander, when it comes to monetary policy, there is no escaping the need for situation-dependent judgment. On this point Colander is probably preaching to the converted as his is a view held by many central bankers and influential monetary theorists such as Goodhart (1988) and Blinder (1998).

On the matter of the training of economists Colander is particularly critical of the emphasis on modelling (mathematical skills) to the exclusion of institutional detail. From his perspective, (p. 66) the major flaw in the current training of economists is their lack of exposure to institutional detail (there are no incentives and numerous disincentives to seek such exposure. In this respect Colander’s anecdotes about Paul Volker’s experience with academics are particularly telling). Nevertheless, when economists move into the real world of government or the private sector it is possible for them to become socially useful. In Colander’s view it takes two months for a smart graduate to realise that most of what was learnt at graduate school is irrelevant and another two years to pick up the necessary institutional detail. Overall this dichotomy in their education is inefficient and in any event leaves academic economists with little exposure to institutional detail.
Colander’s proposed solution is to introduce students to the art of economics at graduate school.

Most American academic economists apparently agreed with Colander’s assessment of the situation in the profession in the late 1980s and a Commission on Graduate Education in Economics (COGEE) was established and published its report in the *American Economic Review* in 1991. The recommendations made by the commission are evaluated by Colander in Chapter 10 titled: The Sounds of Silence: The Profession’s Response to the COGEE Report. As the chapter’s title suggests the response of the academic community to the report has been unfavourable from Colander’s perspective. In the top US graduate Schools there has been no change. Colander is concerned because technical applied work is so difficult — you need to be a good generalist and a good formalist to do it — but the selection process is biased in favour of formalists and against generalists. Consequently, he is of the opinion that the ‘common sense anchor’ provided by the generalists is being lost to the profession.

The remaining chapters in the book apply Colander’s perspective on the art of economics to a wide range of topics. Interesting examples include advice about the teaching of undergraduate macroeconomics, *Telling Better Stories in Introductory Macro*; advice on how to survive as an unorthodox academic economist in a world of US academe, *Surviving as a Slightly out of Sync Economist*; to speculation about the future of the profession, *New Millennium Economics in 2050*. His discussion of the future of the profession makes it clear that he is not against mathematics and models *per se* — just their application without an understanding of the relevant institutional and operational detail. Academic economics will also find his views on the future structure of the US market for the education of economists particularly interesting.

In assessing Colander’s thesis the obvious response is that it misses the point. The work of academic economists and the training of economists is already efficient; it is simply a reflection of comparative advantage. Academic economists have a comparative advantage in model construction while applied economists in government and the private sector have a comparative advantage in institutional detail. The additional two years and two months it takes to produce an applied economist is the lowest cost option. Colander (p. 68) rejects this view and stresses he is not arguing that we should ignore comparative advantage by having academics teach operational detail. He is arguing for recognition by academic economists that operational detail is important and that it has consequences for the development of relevant economic models. That Colander does have a point here can be illustrated by some recent work in monetary theory where neglect of operational detail has produced the sort of consequences he would predict.

There has been much work in recent years on the implications for monetary policy of the evolution of the monetary system to greater use of electronic means of payment. Examination of this issue reveals that some of the models employed are inappropriate and overlook simple operational facts. For example, some early attempts employed an Arrow-Debreu model but that set up was never intended for
analysis of policy issues in general or the evolution of the payments system in particular (Buiter, 1999). As Colander reports, when questioned about the policy relevance of his analysis Debreu replied that it has none. That has not deterred some academic economists from ignoring that advice. Colander’s point is that many of the below-age 45 set of highly technical academic economists are not always as careful as a Debreu or a Lucas when it comes to promoting the policy relevance of their work. Consequently important policy questions are sometimes forced into a non-operational straight jacket with the inevitable result that any purported policy advice is simply ignored. This state of affairs could be avoided by a return to a focus on the art of economics at graduate school.

To sum up, both the COGEE report and Colander are asking only for a change at the margin — some exposure at graduate level to the importance of operational detail. As very few academic economists will make significant theoretical contributions Colander has a point that exposure to the art of economics at graduate school will yield a net benefit. Most academic economists will be involved with policy issues during their careers and even the theorists may find inspiration from real problems they would otherwise have overlooked. A notable recent example is Woodford (2001) who has produced some useful analysis after exposure to the operating procedures on the central banks of Australia, Canada, and New Zealand. Woodford’s paper neatly illustrates Colander’s stress on the importance for theoretical work of intricate operational detail. Whatever view one might take of Colander’s thesis he continues to ask important questions and offer constructive advice about the direction the profession is taking. In the main he undertakes the task without getting bogged down in esoteric questions of methodology or the philosophy of science. Those who are interested in the health and future direction of the profession will find these essays a stimulating, if somewhat provocative, read.

References


Colin Rogers is Dean of the School of Economics at the University of Adelaide.
NON-AGENDA

With the view of causing an increase to take place in the mass of national wealth, or with a view to increase of the means either of subsistence or enjoyment, without some special reason, the general rule is, that nothing ought to be done or attempted by government. The motto, or watchword of government, on these occasions, ought to be — Be quiet...Whatever measures, therefore, cannot be justified as exceptions to that rule, may be considered as non-agenda on the part of government.

—Jeremy Bentham (c.1801)

The Economics of Pitch Invasions

Ron Duncan

Invasions of sports fields by large numbers of ‘fans’ seem to have become more common in recent times. The crowds are mostly amiable but sometimes there is aggressive behaviour towards the players and umpires. As well as posing some threat to players, umpires and security personnel, pitch invasions occurring during the course of an event spoil the entertainment for other fans. These invasions of cricket ovals or football fields should be distinguished from transgressions by individual spectators who mostly appear to be driven by exhibitionist tendencies. I argue that the two behaviours are quite different and should be handled differently.

There is a form of self-enforcement or internal control within crowds that aims at minimising disruptions, so that a disturbance created by a member of the crowd does not spoil the enjoyment of listening to a concert or watching a movie or a sporting event. The disturbance creating by talking or unwrapping sweets during a performance draws the ire of the crowd, which is usually enough to control it. In a sense, therefore, there is no free-riding by individual members of a crowd.

There are external controls of crowd behaviour as well as the internal controls. There are security people at sports events or ushers in theatres; also there are legal sanctions such as fines for extreme misbehaviour. While the internal and external forms of control appear to be effective with respect to individuals’ misbehaviour, they do not appear to be effective against mass...

Ron Duncan is Professor and Executive Director of the Australia Pacific School of Economics and Management at the Australian National University.
misbehaviour such as pitch invasions. To understand why, we need to understand the different dynamics or economics at work.

An individual running onto a pitch or making a noise in a theatre has weighed up the benefits and the costs and decided that the cost of the internal or external sanctions is more than compensated by the benefits (exhibitionist pleasure, bets with friends, etc). Large group misbehaviour within a crowd leading to a pitch invasion, or what might be called a ‘rebellion’, has different economic underpinnings. The expected cost to an individual in a rebellion will be lower the larger the group, as the expected probability of being sanctioned falls with the size of the rebellion. Therefore, the benefit to the individual of their anti-crowd behaviour can easily outweigh the cost of participating. Moreover, as with any rebellion there must be a leader or leaders of a pitch invasion. And it is in the interest of the leader(s) to increase the size of the invasion force so as to reduce the expected cost to each individual, including themselves. The internal mechanism of crowd control is not effective against a rebellion; again for the reason that the individual cost of any sanction is unnoticed and therefore irrelevant. Moreover, the personal cost to well-behaved members of the crowd of trying to impose controls over the unruly mob is likely to be very large.

Because the dynamics of individual and mass misbehaviour are different, the approaches towards control of individual misbehaviour at sporting events and pitch invasions should be different. Individual misbehaviour is best controlled by sanctions (such as fines). As we have seen in Australia in recent years, the incidence of individuals running onto the pitch has been greatly reduced by the use of heavy fines. Some may argue that sanctions do not affect behaviour. But one only has to look at the difference in the behaviour between ice-hockey players, where a punch may lead to two minutes in the penalty box, and American (gridiron) football, where a punch will very likely lead to the loss of several months of income, to see that sanctions can be very effective. A game of ice-hockey is often a sequence of fights between players, interrupted by some ice hockey. One rarely sees a punch thrown in a game of gridiron.

In the case of pitch invasions or mass misbehaviour, ex post sanctions are unlikely even to be applied, as generally seen unless someone is actually assaulted. Therefore, identification of the leaders or inciters of crowd misbehaviour is likely to be the best way of reducing the incidence of pitch invasions. Thus, recent police and stadium action in the United Kingdom and Australia of identifying inciters of misbehaviour and prohibiting them from entry to the sports arena seems to be the most appropriate form of control. Perhaps they should also be charged the total cost of the damage caused by the invasion.
How Should Higher Education Be Funded?

Andrew Norton

During the 2001 election campaign support for extra higher education funding came not just from the predictable higher education interest groups, but also from Rupert Murdoch and the Business Council of Australia’s President, John Schubert. Including universities in their arguments about education spending, the two business leaders made similar points, with Mr Murdoch saying that ‘the brains, skills and entrepreneurial spirit of its citizens are a country’s most precious and powerful asset’ (Murdoch, 2001:2) and Dr Schubert that ‘the skills, ingenuity and know-how of our people will be the primary determinant of social, political and economic success’ (Schubert, 2001:5). Added educational investment is needed to ensure we achieve these human capital goals.

While, as will be explained, a good case can be made for additional expenditure, both men missed several steps of the argument. In education as in other investments, more inputs do not lead inexorably to commensurately greater outputs. We need to consider questions of allocative efficiency. Even if more money became available, are we able to direct the right levels of investment, to the right people, at the right time and in the right place?

The reasons why the answer to these questions is ‘no’ are outlined in this article. There is evidence that the current system both systematically under-invests in higher education and mis-invests too much of the capital it does possess. Replacing the current system of centralised control with improved market mechanisms would enhance Australia’s higher education performance.

The Status Quo

At present, we have a top-down approach to spending on higher education. The Commonwealth sets the total number of Australian undergraduate student places it will fund, and divides these places between the existing universities, except for Bond, which enrols only private full-fee paying undergraduate students. The main factor driving the total number of places is the Commonwealth’s budget situation, and the main factor driving the division of places between universities is history, though political factors, unmet demand, and tendering have also played a role over recent years. The Commonwealth lets universities enrol extra students above the number it sets, either as Higher Education Contribution Scheme (HECS) students or as full-fee paying students. However, this does not dramatically affect total student numbers, as the ‘over-enrolled’ HECS students are financed at only about

Andrew Norton is a Research Fellow at The Centre for Independent Studies, and was Higher Education Adviser to Dr David Kemp when he was the federal Minister for Education.
a quarter of average full-funding, and the Australian full-fee paying students are capped in number and cannot access government loans. Considerations of allocative efficiency, as an economist may think of them, are relevant only at the margins of Commonwealth policy, sometimes affecting the small number of new places, but not existing positions.

The Commonwealth can influence the distribution of students between faculties through the annual ‘profiles’ process of discussions between it and the universities. In practice, however, universities are generally given freedom to decide place distribution for themselves. Students are also free to apply for whichever university and course they choose; the Commonwealth has no power in this regard. Consequently, while heavily constrained by Commonwealth parameters, universities and students determine the final allocation of places.

Commonwealth control is greatest over how much is invested in each student. It sets a subsidy per student, but the actual average subsidy is less, because most universities have some so called ‘over-enrolments’ that push down the average, because they are funded at a low rate. On Australian Vice-Chancellors’ Committee (AVCC, 2001a) figures, in 2000 average ‘full’ funding was $12,003 per student, and the actual average $11,460. While universities are free to spend more than their subsidy on these students, they are forbidden from charging them fees (HECS goes to the government, not the universities). Information on actual per student spending, as opposed to subsidy, is impossible to obtain, but given the limited financial resources available to Australia’s universities it is unlikely to be significantly higher than average Commonwealth funding levels.

### Identifying Allocative Efficiency

We cannot comprehensively and conclusively say what allocative efficiency looks like in higher education. There is no community agreement on the goals of higher education; in particular there is a divide between the pursuit of knowledge for its own sake and an instrumental view of universities as training centres for the professions. Both views are found in university applicant surveys, with interest in exploring the field of knowledge and opportunities for interesting and rewarding careers both rated as strong or very strong influences on course choice by 85 per cent or more of respondents (James, Baldwin and McInnis, 1999:19). Universities are also commonly expected to provide community service through academic advice and comment on matters of public interest, to support local regions, and advance ‘equity’. Neither the need for these public benefit activities nor universities’ success in providing them is easy to quantify. However, in areas such as maximising financial returns to education, satisfying student study preferences, and meeting labour force requirements, there is data suggesting under- and mis-investment in higher education.

Research consistently shows that there are good private returns to higher education, with a recent study showing an average private rate of return of 14.5 per cent, and some disciplines with average returns of around 19 per cent (Borland, 2001:Table 4). Prospective students ought to be attracted to these high
returns and invest strongly in university education. In the largely unregulated area of overseas students, that prediction is confirmed, with full-fee paying overseas students revenue increasing by around 75 per cent between 1996 and 2000 (Nelson, 2001:Figure 1.22). In the highly regulated area of educating Australian undergraduate students, however, investment (as measured by the university operating grant) increased by only 1.6 per cent in the same time period (AVCC, 2001a).

This low figure is due partly to fewer people attending university than probably would have been the case without government restrictions. Throughout this period there has been ‘unmet demand’ — that is people who had applied to attend university but did not receive an offer. For 2001, the Australian Vice-Chancellors’ Committee estimated that there were between 9,400 and 15,250 such persons (AVCC, 2001b). This is a conservative estimate, removing from the applications data students who had applied in more than one state or did not realistically satisfy entry requirements. If it were not for the Commonwealth’s limits on the number of places is likely that universities would have enrolled some or all of these students.

Aside from those missing out entirely, we can hypothesise that students who did enrol would, if given the option, invest more to fix acknowledged problems in the current system, such as a lack of individual attention. The Course Experience Questionnaire (CEQ) asks completing students whether they agree that staff put a lot of time into commenting on their work. Only nine per cent strongly agree, and a further 25 per cent agree, though less strongly. The rest ranged from a neutral response to strongly disagreeing (Graduate Careers Council of Australia, 2001:7). These low levels of satisfaction raise educational quality issues. A major literature survey on developing intellectual ability found that ‘substantial evidence exists to suggest that interactions with major socialising agents (faculty and peers) are, in fact, significantly linked to development of general cognitive skills during college’ (Pascarella and Terenzini, 1991:149). This suggests that student-staff ratios should be reduced. In fact over the last decade student-staff ratios escalated from an average of around 13 to one in 1990 to about 19 to one in 2000 (AVCC, 2001c). A greater investment in higher education could remedy these problems, but current regulation forbids students from making this investment.

There is also evidence of false starts in higher education, with students enrolling in the ‘wrong’ course. Nearly a third of first year undergraduates in a 1999 survey were not enrolled in their first preference course (McInnis, James and Hartley, 2000a:15). Persistence of the original preference contributes to around 17 per cent of students changing faculty or course in their first or second year (McInnis et al., 2000b:30). Taking 17 per cent of the commencing students and multiplying that by the average government subsidy (net of HECS) arrives at a taxpayer cost of around $365,000,000 a year from course changes. The figure should be discounted for people who retain credit in their new course for work they have done in their old course, but enlarged for those who drop out entirely for course related reasons. A precise figure for course change costs is difficult to calculate, but it is almost certainly in the hundreds of millions of dollars annually.
Though students switching to their preferred courses are better off in the long term, they incur additional HECS liabilities, direct expenses, and for full-time students opportunity costs from being out of the workforce while studying.

We also see very high rates of non-completion in some disciplines. In 1999, less than 60 per cent of students who had enrolled in Arts, Engineering and Science in 1992 had completed their degrees, compared to over 70 per cent in Education, Health, and Law (Martin, MacLachlan and Karmel, 2001:8). While non-completion does not mean no benefit was derived from enrolment, high attrition does suggest the investment in these students could have earned better returns elsewhere. A study of the responsiveness of universities to student demand found that there was evidence of it, but there was oversupply in Agriculture, Engineering and Science (two of the poor completers) and excess demand in Architecture, Health, Law, and Veterinary Science (Li, Karmel and MacLachlan, 2001:21). Correcting these imbalances may reduce the number of weakly committed students, and improve completion rates.

To some extent, these allocative problems spill over into the labour market. According to 2001 skills shortages data, there are wide ranging shortages of health professionals, with national shortages in fifteen specialities. In engineering, there is a national shortage of electronics engineers and regional shortages of civil and electrical engineers (Department of Employment, Workplace Relations, and Small Business, 2001). While the labour market is difficult to predict, the co-existence of excess demand amongst prospective students wanting to enter health professions, and shortages of such people in the labour market, suggests a failure to allocate educational resources efficiently. One reason for this is that health courses are relatively expensive to run, but universities have discretion about how they reach their government target number of students. It makes sense for the universities to meet their target with cheap courses, such as humanities and business, rather than health, even though this is sub-optimal for the students, society and the economy.

**A Market Alternative**

Overall, then, there is considerable evidence of allocative inefficiency in the current centrally controlled system. A market system, such as that unsuccessfully proposed by the then Education Minister, Dr David Kemp, in 1999 would, in theory, do a much better job. Kemp proposed to abolish limits on place numbers; all students accepted by a university would be entitled to a subsidy and a loan. He also proposed abolishing the Commonwealth’s role in allocating student places between universities; instead universities would be paid for the number of students they enrolled, through a mix of Commonwealth subsidy and student fees, deferrable with an income contingent loan. In order to fill their places, universities would need to be more responsive to student demand. In effect, students would have the role the Commonwealth has now (Kemp, 1999).
Information Problems?

Rather than directly defend the current system, commentators tend to focus on the perceived deficiencies of the market alternative, and in particular the information problems of a large percentage of prospective university students, recent school leavers. Clive Hamilton (2001:12) observes that eighteen year olds ‘frequently have no good idea of what is in their own interests, let alone how their education can contribute to society’s interests’. Gabrielle Baldwin and Richard James (2000:142) note that much of what occurs in a university education is intangible and non-observable, and outcomes are long-term. Harry Clarke (2000) argues that higher education involves an ‘information asymmetry’, as academic producers know more than student consumers. This is particularly serious for new courses without alumni or market reputation. Empirical evidence supports the view that prospective students experience a knowledge problem. A survey of school leavers showed that many do not class themselves as very well-informed, even about important matters such as academic services for students or employment outcomes (James, Baldwin and McInnis, 1999:27).

On these commentators’ view, a market system would not improve things greatly. Students’ ignorance contributes to the existing misallocation of resources and would exacerbate it if they were given more influence. This, however, reflects an overly pessimistic view of the state of student knowledge, and what can be done to improve it.

Studies of Year 12 students show that they often display consistent fields of interest. People with artistic interests tend to apply for visual arts and music courses; people with social interests apply for childcare, community service, and health studies; people with investigative interests apply for engineering, computing, and applied science; and so on (Harvey-Beavis and Elsworth, 1998:79). These consistent interests are displayed in the way they fill in their application preferences, with applicants in many fields putting down multiple similar courses rather than applying for a variety of different types of courses (Harvey-Beavis and Elsworth, 1998:51). There is no evidence that forcing students into courses that do not interest them will produce a public benefit, as seems to presumed by those sceptical of eighteen-year olds’ ability to make higher education choices. It is more likely that ignoring student preferences will, as described above, end in course changes or dropping out. Under a market system of higher education, in which no university was guaranteed its student numbers and new universities could enter the market, there would be stronger incentive to match student demand.

While students typically know their general interests, this does not straightforwardly lead them to a particular course. They need to know more about the content and outcomes of courses. A considerable amount of information about outcomes is produced, and it is becoming increasingly accessible. All Australian universities participate in the Graduate Careers Council of Australia’s Graduate Destination Survey (GDS), which reports on employability and starting salaries, useful information for students enrolling for vocational reasons. This information
is now available free in an easy to understand format from the Graduate Careers Council’s Gradsonline website (http://www.gradlink.edu.au) and from the Commonwealth Education Department’s website in its Which Course? Which University? section (http://www.detya.gov.au/tenfields/), and via a starring system in the commercial Good Universities Guide. The GDS is conducted around four months after completion, and while longer-term information is more difficult to obtain, the Australian Bureau of Statistics (ABS) publishes data on unemployment by field of study (ABS, 1999:41), and some statistics on proportions of graduates working in jobs unrelated to their original field of study (ABS, 2001:112). Academic studies provide information on the average long-term returns graduates received on their higher education investment (Borland, 2001).

While in a flexible labour market there is no guarantee that historical experience is an accurate guide to the future, the available information on labour market performance is nevertheless useful for prospective students. The fact that some fields of study have for many years had poor employment rates, poor salaries, and high proportions of graduates in jobs working in areas unrelated to their degrees, is information that ought to be better known than it is. It would be useful if publications such as the Good Universities Guide included more details on these matters.

All Australian universities also participate in the CEQ, which gives prospective students useful information about how very recent students rate their educational experience. There are considerable differences between universities, and these are also reported via a starring system in the Good Universities Guide. The research intensive universities tend to do particularly badly, while small and regional campuses do better than average — important information for those unsure of their academic abilities.

While information is available, there are obstacles to its effective use. The universities themselves use it only sparingly, perhaps because the results tend to be mediocre. Only the commercial Good Universities Guide has a strong incentive to publicise the information. While the Good Universities Guide can assist prospective students in narrowing their options, it still leaves an information asymmetry on the detail of particular courses. While it is true that producers know more about their own courses than student-consumers, this does not mean that they should decide how educational resources should be distributed. They are ignorant of prospective students’ fields of interest. They are also likely to be under-informed about what other universities have to offer, and have a vested interest in promoting their own course.

We need people who can get around the information problems prospective students face, and avoid the producers’ conflict of interest. An education broker would put the interests of the prospective student first, but bring to the task knowledge of courses usually only held by the university. Existing research on what influences school leavers in their higher education choice shows that personal sources of information, such as material given to them by careers teachers, and what they learn at university open days, rank highly among the strong influences (James, Baldwin and McInnis, 1999:15). A broker should
replicate this interpersonal element, helping to win trust. The difference between an education broker and a school careers adviser is that the broker would specialise in one or a small number of fields, so as to maintain the requisite levels of expertise. Sophisticated broker services could use aptitude testing to help determine what kind of course and institution would be best, and advise on various pitfalls of first year life, such as not adapting to new learning styles, social isolation, and time allocation. In relatively recent times broker services have developed for the vocational training market, and have generally been successful. One study of their operation in regional Australia endorsed public support of them, on equity grounds for their value in bringing low income people to education, and for their role in promoting efficiency through better matching training and provision (Kilpatrick and Bound, 2001:66). In the training sector, there are a mix of profit and not-for-profit broking services. Whatever their funding source, brokers are likely to be money well spent through reducing costs to students and the taxpayer from wrong choices.

Where Should the Money Come From?

As suggested at the start, it is common ground that more money ought to be invested in higher education. There is evidence that added investment translates into improved financial returns. In the United States, one recent study found that each $1,000 increase in tuition expenditure was associated with increases in male earnings of about two per cent (Fitzgerald and Burns, 2000:33). There is not a single theory explaining this, but improved cognitive development from employing better staff and increasing interaction between staff and students fits with the educational research (Pascarella and Terenzini, 1991). In the Australian context, research is finding good returns to cognitive skill (Pappas, 2001), so education that develops such skill is likely to be rewarded in the labour market.

The point of dispute is which mechanisms ought to be used for increasing investment. Most lobby groups and commentators call for increased public investment, distributed through the current system (either expressly, or by implication in not specifying any alternative). Even if the current system was adapted to improve use of student course preferences, such as introducing a voucher system, this strategy is still likely to produce sub-optimal results. The most immediate problem is that public investment is unlikely to be forthcoming, as there are strong electoral reasons why the Commonwealth is not going to give higher education funding priority (Norton, 2001). But even supposing that the money was forthcoming, the existing incentive structure does not encourage investment in student learning.

With the current system, there is an incentive for universities to invest in research and not teaching. For example, the Australian Research Council (ARC) awards its grants on the basis of excellence, and success in ARC grants in turn affects the general research subsidy each university receives. In teaching, there are flat subsidies, neither rewarding excellence nor penalising mediocrity. The effects of this incentive system are very evident in university practices. A 1999
survey of academics found that only 44 per cent of early career academics had received training at the start of their career, and that figure dropped to around 30 per cent for mid and late career academics (McInnis, 1999:32). Instead the standard qualification needed for a career as a university teacher is a research degree, the PhD. The survey also asked academics about their time use, and compared its results with a 1993 survey. It found that—despite the intervening increase in student-staff ratios—the amount of time spent weekly on teaching and teaching related activities went down by 1.3 hours and the amount of time spent on research went up by 1.1 hours (McInnis, 1999:26). This is the opposite of what you would expect if universities gave students first priority, but exactly what you would expect if you believe that universities respond to incentives. On this theory of incentives, gains from additional public investment in the existing higher education system would be lower than the same amount coming from private student investors in a market system. US research finds that greater reliance on private sources of income has a positive impact on teaching quality, which is consistent with my Australian analysis (Brown, 2001).

The Commonwealth could try to create incentives within the existing system by rewarding performance. It is already developing the Graduate Skills Assessment (GSA) test, a measure of generic skills, such as written communication, critical thinking, problem solving and interpersonal understandings (Australian Council for Educational Research, 2001). This could be conducted at entry and exit level to find some measure of value adding performance. But while these generic skills are good in themselves, such performance measures risk perverse incentives. Universities may excessively teach to the test, rather than focusing on other skills or specific content. It may discourage forms of innovation that have unknown effects on the testable skills. The GSA is likely to provide valuable information, but it is better given to prospective students as part of the information used to make a rounded decision.

A system of Commonwealth control of university finance also prevents students making sensible trade-offs. It makes sense for highly intelligent and ambitious people to invest more in their education than those who are less able or less committed to the labour market. Yet the current system forces everyone into an average funding model, investing too little in some people, and possibly too much in others. There is currently no price differentiation between those who want the full campus experience, and those content with minimal interaction. Mature age students, particularly, may have neither the time nor the desire to mix with other students, while for younger students this can be important for socialising and learning. Letting universities control prices would increase capacity to offer different types of experiences to different groups. We already see this in the deregulated postgraduate student market, with bargain Internet MBAs available for as little as $10,800 a year, ranging up to $42,000 at a high service, high prestige institution (Ashenden and Milligan, 2001:156-169).
Conclusion

The current education federal Education Minister, Dr Brendan Nelson, has said that the status quo ‘is not sustainable’. He has also said that further significant public funding increases will not be forthcoming (Nelson, 2002). There is little room for allocative inefficiency in this financially constrained environment. Introducing a more market oriented system would create new incentives to allocate limited resources in an efficient manner.

References


Australian Vice-Chancellors’ Committee (2001a), ‘AVCC Funding Table 3: Commonwealth Resources (base operating grant) Available to Higher Education Institutions per Planned and Actual EFTSU, 1983-2003’, available [www.avcc.edu.au](http://www.avcc.edu.au)


Australian Vice-Chancellors’ Committee (2001c), ‘Actual Student:Staff Ratios by AVCC Institution and Academic Organisational Unit’, available [www.avcc.edu.au](http://www.avcc.edu.au)


*The arguments in this article are elaborated in a forthcoming CIS book. The author is grateful for helpful comments received from two anonymous referees on an earlier version of this paper.*