Incomes and Assets of Older Australians: 
Trends and Policy Implications

Ann Harding, Anthony King and Simon Kelly

This paper examines trends in the wealth and incomes of older Australians over the 11 years from 1985-86 to 1996-97. The primary data sources for this analysis are the ABS 1986 Income Survey and 1997-98 Survey of Income and Housing Costs ‘confidentialised’ unit record files. These surveys provide very comprehensive national snapshots of the income, socio-demographic and other characteristics of a representative sample of Australians.

The ABS defines income as ‘cash receipts that are regular and recurring’ and includes income sources such as wages and salaries, profit and loss from own business, property income, government cash transfers and private cash transfers (such as superannuation and child support) are all included. Receipts that are excluded from income because they are not regular and recurring include inheritances, maturity payments on life insurance policies and capital gains and losses. Imputed rent is not calculated for home owners. Annual cash income is the income measure used.

While income is a flow, wealth is a stock. Both are important when considering the economic position of older Australians. Two older Australians with similar incomes might be in very different positions if one owns no assets but the other has substantial assets to draw upon if and when they are in need.

The ABS surveys provide details of only one type of wealth — the respondent’s estimate of the value of their home and the amount of mortgage outstanding on that home. The surveys do not directly provide estimates of the value of other wealth sources, but instead report the income received from such wealth sources (such as interest, rent and dividends). This information has been used by NATSEM to impute estimated wealth holdings for each of the families contained within the survey. For 1986 this utilised the methodology mapped out by Bækgaard (1998) with some later minor adjustments. For 1996-97 the methodology is described in Kelly (2001). The wealth sources included in this study comprise own home, shares, cash, investment properties, own business and accumulated value of superannuation. Other sources, such as consumer durables and cars, are not included. The discounted future value of the government age pension is also not included within our estimates of private wealth holdings.

It must be emphasised that imputing the estimated value of wealth holdings using such capitation methods, whilst a technique frequently employed by researchers in this field (Dilnot, 1990), is nonetheless subject to some degree of

---

*Ann Harding is Professor of Applied Economics and Social Policy at the University of Canberra and the inaugural Director of NATSEM. Anthony King is a Principal Research Fellow at NATSEM. Simon Kelly is the current holder of the NATSEM PhD Scholarship.*
uncertainty. In addition, there may be issues of comparability between the 1986 and 1997-98 Income Surveys. For example, in 1986 the definition of dependent children is somewhat narrower than in 1997-98. In 1997-98 negative business and investment incomes were left untouched by the ABS while, in 1986 they were reset to zero. At this stage we have not amended the data to attempt to standardise for such differences. In addition, some issues have been raised about the comparability of the ABS income surveys over time (given the change in methodology from the mid 1990s surveys onwards), and in a joint project the ABS and the Social Policy Research Centre at the University of NSW are currently examining these issues. It is also important to note that the ABS only surveys those living in private dwellings so that the small proportion of aged persons living, for example, in nursing homes, are not included within the scope of our estimates. Finally, both sampling and non-sampling error affect all sample survey results.

There are a number of possible ways of looking at the economic wellbeing of older Australians. Here we are looking at families categorised by the age of their head. Our ‘older Australian’ group thus includes a couple where the husband is aged 65 and the wife is aged 60, but may exclude Australians aged over 65 years who live in families where a younger spouse has been categorised as the head. Our families comprise either a single person, a sole parent living with their dependent children, or a couple with or without dependent children.

The focus in this article on trends in the wealth holdings of older Australians complements other recent research into the changing economic circumstances of this group. Whiteford and Bond (2000) examined trends in incomes from 1986 to 1995-96, with consideration too of the role of non-cash benefits, indirect taxes, housing wealth, and the assets of pensioners (see also King, Bækgaard and Harding, 2001). Recent comprehensive estimates of the wealth of older Australians (Bacon 1998) have, however, been for a single point in time. Here, we present estimates of trends in wealth holdings among older Australians.

In considering the policy implications of the trends that emerge, the context is one of considerable policy activity in the area of retirement incomes over the past two decades (see Bateman, 1999; and King, Walker and Harding, 2001, for summaries). The Australian publicly funded pension scheme remains unique internationally, with our age pension being a non-contributory flat-rate means-tested system, with eligibility governed by age and residency criteria. Some 80 per cent of Australians who qualify by age receive either all or a portion of the full age pension.

However, in preparing for more rapid population ageing over the next two decades, successive Australian governments have made several changes to the retirement income system. These changes — which include a shift toward greater self-provision via the compulsory Superannuation Guarantee levied upon employers and, for women, an increase in the age of eligibility of the public pension — are in line with policy shifts observed in many other developed countries.
The dominance of the public age pension in Australia has until now ensured greater uniformity among the incomes of older persons than is common overseas. An increase in the diversity of the incomes and wealth holdings of older Australians can be expected in the future, as the impact of more comprehensive superannuation in the post-war period bears fruit — and such growing diversity is already apparent over the past decade.

**Changes in Wealth**

Our estimates suggest a 62 per cent increase in the real value of total Australian household wealth from the mid 1980s to the late 1990s. Expressed in 1998 dollars, this is an increase from $1058 billion in 1986 rising to $1712 billion by 1997. How much of this wealth is owned by older Australians? Has their share of wealth increased? Has its composition changed, or are the same types of asset as important now as they were in the mid 1980s? Has the pattern of wealth holdings among older Australians changed?

**Figure 1: Estimated Shares of Total Net Australian Household Wealth by Age of Family Reference Person: 1985-86 and 1996-97**

![Figure 1: Estimated Shares of Total Net Australian Household Wealth by Age of Family Reference Person: 1985-86 and 1996-97](image)

Data source: NATSEM simulations based on unit record data from ABS 1986 Income Distribution Survey and ABS 1997-98 Survey of Incomes and Housing. For the detailed tables upon which this and other figures are based, see Harding, King and Kelly, 2002.

**Shares of Total Wealth**

The increasing share of wealth held by older Australians is clearly evident in Figure 1. The distribution of wealth across the population has shifted markedly toward older Australians since the mid 1980s. The share held by those aged 65 or over increased from 17 per cent to 27 per cent between 1986 and 1997. Conversely, younger Australians aged less than 45 years held a declining share of total wealth.
Now, we know that at the same time the share of the Australian population aged 65 or more has also been increasing. So, how much of this greater share of wealth held by older Australians simply reflects higher population numbers in these age groups? Looking at average wealth per adult provides a much clearer picture and this is done in Figure 2.

**Figure 2: Estimated Net Average Wealth per Adult by Age of Family Reference Person: 1985-86 and 1996-97 (June 1998 dollars)**

![Figure 2](image)

Note: To derive all ‘per adult’ estimates in this paper, the total wealth of couples has been divided equally between the two partners. Note that all estimates are in June 1998 dollars, with adjustment by the CPI.


Figure 2 shows a dramatic change in the picture since the mid 1980s. For those living in families with heads under 45 years old, average wealth per adult actually fell over the period. For those aged 45 years and over, average wealth per adult rose. This increase in wealth was particularly marked for older Australians aged 65 years and over. Average wealth for the 65-74 years age group increased by 115 per cent — from an estimated $103,000 in 1986 to $221,000 by 1997. The increase for the 75 and over group was not as high, but still a very substantial 61 per cent — from $112,000 in 1986 to $180,000 by 1997. And remember that these figures are in constant June 1998 dollars, so they reflect real increases in wealth and not just the effects of inflation.

Understanding the sort of changing pattern shown in Figure 2 can be quite complicated because it is necessary to disentangle so-called ‘age’, ‘period’ and ‘cohort’ effects. The ‘age’ effects are those where we see a distinctive pattern to wealth accumulation over the life course — increasing wealth with age until retirement when there is typically some running down of wealth. This broad pattern is exhibited in Figure 2 — though there is a notable difference between 1986 and 1997 in the apparent degree of running down of assets in older age.
‘Period’ effects are those that reflect different conditions prevailing at different times. For example, the returns on investments, taxation regimes, and the ease of access to home-ownership vary over time. People’s wealth accumulation will clearly be affected by the conditions prevailing over different stages of their life. Finally, ‘cohort’ effects refer to the possibility of cohorts of the population behaving differently. Thus, people aged in their fifties in the 1970s may well have behaved differently to people aged in their fifties in the 1950s. And we know that there are many such differences — in, for example, family formation, labour force activity and so forth.

The importance of period and cohort effects in the picture of the changing wealth of older Australians is very evident in Figure 2, which gives the impression of a wave moving through the age profile of wealth holdings. In 1986, the peak wealth holding was for the 55-64 year age group. Eleven years later, the peak is observed for the 65-74 year age group. Similarly, the age at which average wealth starts to decline has advanced 10 years over the 11-year period. There is thus a very strong suggestion of important period and cohort effects in the wealth holdings of older Australians. Really, this simply confirms the common sense view that the wealth of older Australians will be largely determined by their behaviour and fortunes earlier in life.

What Figure 2 also shows is that there are significant differences in the behaviour and fortunes of population cohorts. The peak wealth holding is seen for those aged 65-74 years in 1997. This population group was born between 1923 and 1932 and their prime wealth accumulation years coincided with the long post-World War II economic boom. In 1997 76 per cent of this cohort owned their homes outright, with the majority thus benefiting from the boom in house prices over the previous decade. The oldest cohort — those aged more than 75 years in 1996 — were born before 1922. The 1930s Depression and World War II coincided with their early adult lives, and they show notably lower wealth holdings in older age than the economically more fortunate cohort that followed.

The cohorts coming after the ‘peak’ 65-74 year old cohort in 1997 also do not look quite so fortunate, with Figure 2 in particular showing real declines over recent years in average wealth for families with heads aged less than 45 years. Further investigation indicated that this was due to falling home ownership rates and the growing value of mortgages for younger cohorts, rather than to declines in the average price of homes. For example, the proportion of families headed by a 25 to 34 year old who either owned or were purchasing their own home fell from 46 per cent in 1986 to 35 per cent in 1997. For 35 to 44 year olds, the comparable fall was 72 to 62 per cent over the 11 years. Allied with these declining home ownership rates were sharp increases in the real value of mortgages. Together these two factors produced falling real home equity values for under 45 year olds. While families headed by 25 to 44 year olds did enjoy increases in the average value of their shares and accumulated superannuation entitlements, these increases were not sufficient to offset their declining home equity. In contrast, older cohorts experienced sharp increases in the value of their superannuation, shares and homes — hence their rising wealth holdings relative to younger groups in the population.
Components of Older Australians’ Wealth

What type of assets do older Australians own? Figure 3 shows that their own home remains the most important asset of older Australians. The net average value per adult of such homes among all Australians aged 65 years and over has increased substantially since the mid 1980s, from about $64,000 in 1986 to $87,000 in 1997 (in 1998 dollars). But despite the continued popularity of the home, other asset classes have shown much more rapid growth among the aged. In 1986 the average superannuation holdings of older Australians were almost non-existent, at less than $3000 per adult. By 1997 this had grown rapidly to $36,000. Here, it needs to be recognised that the fruits of superannuation appear in different forms in older people’s assets, depending on how people take their superannuation benefit. If taken as a lump sum, it may appear, for example, in equities or in interest-bearing deposits. The apparent massive increase in superannuation assets for older Australians will thus partly reflect growing superannuation coverage over the post World War II period, but will mainly reflect the growing likelihood of superannuation benefits being taken as superannuation pensions.

Equities also performed strongly, with the estimated net value of shares directly held by each adult aged 65 years and over quadrupling from about $7000 to $28,000. Interest on cash deposits also rose. The figures below suggest that older Australians may have become more adventurous about investing in the stock market.

Figure 3: Estimated Net Average Value of Various Assets per Adult Aged 65 Years and Over: 1985-86 and 1996-97
Figure 4 shows the effects of these differential growth rates in the various asset classes on the composition of older Australians’ wealth holdings. Direct superannuation assets are now a far more significant component of the total net wealth of older Australians. While superannuation made up only an estimated two per cent of their total wealth in 1986, by 1997 this had risen to almost one-fifth. Equities also became more significant, rising from about 7 to 14 per cent of total wealth over the same period. Despite rising house prices, their own home made up a shrinking proportion of the wealth of older Australians, falling from an estimated 60 per cent of total wealth in 1986 to only 43 per cent by 1997.

Distribution of Older Australians’ Wealth

So far, we have looked at average wealth for older Australians with some distinction only between two age groups — those aged 65-74 years and those aged 75 years and over. How does the picture vary for single males, single females and couples among the population of older Australians? The broad pattern of markedly higher wealth for couples than for singles, and somewhat higher wealth for single males than for single females, has remained consistent over the period. The average wealth level for all three groups grew strongly over the period 1986-1997.

There was, though, some change in the relativities between the wealth levels of couples and single males and females. As a proportion of average wealth for couples, single males’ wealth declined from 64 per cent to 59 per cent over the period from 1986 to 1997. The corresponding figures for single females were 53 per cent falling to 47 per cent. This pattern of changing wealth relativities is partly related to the changing population structure among the population of older
Australians — for example, the extent to which the different groups include people aged 75 years and over. Another important factor, however, is differences in the composition of the asset holdings of couples, single males and single females (Figure 5). The most marked difference in the composition of wealth holdings is the far higher importance of housing wealth in the assets of older single females, compared to single males and couples.

**Figure 5: Asset Classes as a Proportion of Total Net Wealth by Type of Income Unit for Older Australians: 1985-86 and 1996-97**


**Figure 6: Proportion of Total Wealth Received by Older Australians by Quartile of Wealth: 1985-86 and 1996-97**

Another way of looking at wealth distribution is to rank all older Australians by their wealth and then divide them into four equally sized groups — quartiles. Figure 6 suggests that the most profound movement over the 11 years was the strong increase in the share of all older Australians’ wealth received by the top 25 per cent — up from about 67 to 71 per cent. The middle half of older Australians suffered a fall in their share of wealth, while the least wealthy one-quarter saw a slight increase in their wealth share.

Changes in Income

The previous analysis has concentrated upon changes in the net wealth holdings of older Australians. While wealth is an important basis for the economic wellbeing of older Australians, it only provides part of the story. Government benefits — notably the Age Pension and Veterans’ pensions — play a particularly important role, and labour market activity continues to make a contribution for a few. We turn now to a look at the changing incomes of older Australians, which provides a different perspective upon changes in their economic wellbeing over time.

In the following analysis we have looked at annual disposable income (which is gross income minus income tax). Clearly a single older Australian with an income of $20,000 enjoys a different standard of living to an older couple with an income of $20,000. To facilitate more accurate comparison between older Australians living in different types of families, we have used the new OECD equivalence scale to calculate ‘equivalent incomes’. This scale gives a single adult a value of 1, a second adult a value of 0.5 and any dependent children a value of 0.3 each. This scale thus assumes that a couple require 50 per cent more income than a single person to reach the same standard of living. (The amount is not double because of ‘economies of scale’ such as the ability to share housing and heating costs.) The new OECD scale is the scale adopted by Eurostat in its cross-European comparisons.

The equivalent incomes are based on real after-tax incomes to take account of the effect of inflation between 1986 and 1997. As equivalent incomes are a relative rather than absolute measure of income, some form of benchmark is needed. Here, we express the equivalent incomes as an index where the average equivalent income of all those in the 65 years and over age group is set at 1.0.

Overall Incomes

To look at changes in the fortunes of older Australians we have compared them with families with heads in the ‘prime working age groups’ (35 to 54 years). This age group has been chosen because it avoids complications in the comparison that can arise from changes in the behaviour of young people and those in the immediate pre-retirement years. The average incomes of older Australians appear to have increased between 1986 and 1997 by slightly more than for those in the peak working years. The estimated increase in equivalent after-tax annual incomes was four percent for older Australians and 1.6 percent for working age
Australians. (Note that these estimates are after taking out the impact of inflation.) On average, in 1986 the equivalent family incomes of older Australians amounted to 54 per cent of the comparable incomes of families headed by a prime working age Australian. By 1997, this had risen slightly to 55 per cent. This suggests that, on average, older Australians appear to have slightly more than kept pace with rising community incomes.

**Changes by Type of Unit**

Have single older Australians fared better or worse than couples over the 11 years? As Figure 8 shows, even after adjusting for needs, older couples on average enjoy higher equivalent incomes than single older Australians. This partly reflects the younger average age of older couples than older singles, and perhaps a different capacity to invest earlier in their lifetime relative to single people.

As expected, older women have lower average incomes than older men — about 10 percent lower in 1997 for women than for men. The gap between the economic fortunes of single older men and women has also increased over the 11 years, with the average difference in equivalent disposable incomes rising from 10 percent in 1986 to 12 percent in 1997. The faster growth in single male incomes may reflect a preference for growth assets, as well as more rapidly improving superannuation coverage.

**Figure 8: Estimated Mean Equivalent Disposable Incomes of Older Australians by Type of Income Unit: 1985-86 and 1996-97**


**Changes by Type of Unit and Age**

Figure 9 shows changes in average equivalent incomes of older Australians subdivided by both age and family type. Incomes are generally slightly lower for
those aged 75 years and over than for the 65-74 year age group. The difference according to age was much more marked for single males in 1986, but has since been much reduced by a strong increase in the average incomes of single males aged 75 years and over.

**Figure 9: Equivalent Disposable Incomes of Older Australians by Type of Income Unit and Age of Reference Person: 1985-86 and 1996-97**

![Figure 9](chart.png)


**Figure 10: Proportion of Total Income Received by Older Australians by Quartile of Income — Units with Reference Person Aged 65 Years and Over: 1985-86 and 1996-97**

![Figure 10](chart.png)

Distribution of Older Australians’ Incomes

To look at whether higher or lower income older Australians had fared better, we ranked all Australian families headed by an older Australian by their equivalent disposable income and then divided them into four equally sized groups called quartiles. Figure 10 shows the shares of equivalent disposable income received by each quartile of older Australians. The income share of the bottom quartile of Australians fell slightly over the 11 years, while the shares of the top three quartiles increased or remained constant. The increase was particularly strong for the top quartile. As a result, looking just at these older Australian families, the share of all equivalent disposable income received by the bottom quartile fell from 14.1 to 13.1 per cent over the 11 years, while the share received by the top quartile rose from 44.1 to 45.0 per cent. As with the Australian income distribution more generally (Harding and Greenwell, 2002), we have thus seen rising inequality of incomes over this decade among older Australians.

Government Pensions’ Contribution to Incomes of Older Australians

In 1986, government benefits — mainly the Age Pension — accounted for 62 per cent of the after-tax incomes of older Australians. By 1997, this proportion had fallen to 57 per cent. This does not reflect a reduction in the levels of pension payments, which have been effectively indexed in line with earnings, but rather the growing importance of private retirement incomes. As private retirement incomes have increased, the corresponding fall in the contribution of government benefits is compounded by the impact of means-testing of the pension.

Comparing different family types, single females among the population of older Australians have a notably higher dependence on government benefits than the other groups. In 1997, government benefits accounted for 68 per cent of their incomes, compared to the figures of 55 per cent for single males and 49 per cent for couples. This is a direct reflection of the considerably lesser opportunity that these females had to save for private retirement incomes.

The decrease in the role of government benefits in the incomes of older Australians is evident for all three family types, falling from 61 to 55 per cent over the 11 years for single males, from 72 to 68 per cent for single females and from 54 to 49 per cent for older couples. It is, however, particularly marked for couples and single males aged 65-74 years (Figure 11). These two latter groups are those that have most enjoyed the benefits of post-war increases in superannuation coverage.
Incomes and Assets of Older Australians

Policy Implications

After taking out the impact of inflation, the estimated average wealth of Australians aged 65 years and over rose from $106,000 in 1986 to $204,000 in 1997 — about a 90 per cent increase. Older Australians have particularly high home ownership rates, and were thus major beneficiaries of the nominal doubling in house prices over this period. Despite this sharp increase in home equity, even stronger growth was experienced in the net value of shares owned and accumulated superannuation benefits among older Australians. As a result, their own home made up a shrinking proportion of the wealth of older Australians, falling from an estimated 60 per cent of total wealth in 1986 to only 43 per cent by 1997.

This profile of growth in average net wealth was in sharp contrast to the fall in the average wealth holdings of Australian families headed by a 15 to 44 year old over the same time period. For example, in 1998 dollars, the average estimated wealth of Australian families with a head aged 25 to 34 years fell from about $67,000 in 1986 to $58,000 in 1997. Further investigation revealed that this decline was primarily generated by falling home ownership rates allied with rising home mortgages. Even though the average value of shares and accumulated superannuation benefits for these younger generations had risen, this was more than offset by their falling home equity.

The substantial increase in average wealth among older Australians masked varying outcomes for the most and least wealthy within this group. Considering
just older Australians, the share of total wealth held by the wealthiest 25 per cent rose from about 67 to 71 per cent.

Moving away from wealth to look at incomes, on average the incomes of older Australians slightly more than kept pace with those of working age Australians over these 11 years. However, the income share of the most affluent quartile of older Australians increased more quickly than those in the middle two quartiles, while the inflation-adjusted income share of the bottom quartile declined. Government cash benefits declined as a share of the after-tax incomes of the aged, from 62 per cent in 1986 to 57 per cent in 1997, reflecting the growing significance of non-pension income sources.

Overall, the results indicated profound inter-generational shifts in the distribution of wealth, with older Australians recording sharp increases in their net wealth and younger Australians reporting net falls. However, the rosy picture for older Australians is affected by the rising fortunes of the wealthiest among this group, with the remaining older Australians not sharing in the economic good fortunes of the wealthiest to the same extent over the 11 years.

Until recently there has been relatively little inequality among older Australians. In previous decades the vast majority were dependent upon the age pension and accumulated wealth levels were relatively low. As a result of the spread of post-war superannuation and the subsequent introduction of the Superannuation Guarantee, allied with substantial asset price rises, this picture of relative equality is starting to change. Our study reveals that there is growing divergence in the economic fortunes of older Australians.

The growing diversity in the economic wellbeing of older Australians raises a number of challenges for policy makers. Most older Australians remain heavily dependent upon the age pension, with increases in the rate of age pension and/or liberalisation of the income and assets tests continuing to be extremely important in determining the living standards of older Australians. However, a slowly growing proportion of older Australians have substantial assets and private incomes, and mechanisms such as reverse mortgages and the extent to which such Australians can pay for services may assume greater prominence in the policy debate as governments grapple with the challenges of financing the health and care costs of older Australians.

Over the next decade or two the large baby boom cohort born after World War II will start retiring. There is extensive debate about whether and to what extent this will increase the fiscal pressures upon government (Kinnear, 2001; Productivity Commission, 1999). Health care costs are likely to loom large in this debate, as changing medical technologies lead to higher medical and pharmaceutical costs (Walker, Percival and Harding, 2000). The extent to which the retiring baby boomers will be able to help finance their health and care costs in their old age will become an increasingly important issue for governments. It is already clear that many older Australians — particularly women — will have relatively limited incomes and wealth to draw upon (Association of Superannuation Funds of Australia, 1999; Kelly, Percival and Harding 2001; King, 2002). However, our study reveals that the older baby boomers, aged
between about 45 and 52 years in 1997, enjoyed a substantial increase in their average net wealth over the 11 years of this study. Many in this group will have a greater capacity to help finance their future needs in retirement.

As discussed earlier, the generations following the baby boomers have been less successful in building wealth than the boomers were at the same age. We can thus expect growing inter-generational tensions in the future, with the baby boomers supporting higher taxes upon younger generations to help finance their needs in old age, but younger generations questioning whether the boomers can bear a greater share of the financial burden of service provision. Throughout all this, the significance of inheritance is likely to increase. Existing older Australians are wealthier than any previous generation, and the extent to which they run down their assets before death — and who they leave their wealth to — will have important policy implications. The taxation of inheritances is also likely to become a more hotly debated issue.

References


_The authors are grateful to George Rothman and two anonymous referees for helpful comments on earlier versions of this work. Aspects of this study have been financed by Australian Research Council Grant No. A79906127._
Reactive, Not Proactive: Recent Trends in Australian Broadcasting Regulation

Carolyn Lidgerwood

This paper explores how the introduction of digital technologies and the resulting convergence of broadcasting services with other communications services is challenging the current broadcasting regulatory regime in Australia. As the Productivity Commission (2000) recently observed:

With advances in digital technology, broadcasting, telecommunications and the Internet are converging rapidly. They are being fundamentally redefined in terms of what they are, who provides services, and how they are produced and delivered. Broadcasting … is not what it was when the Broadcasting Services Act was introduced in 1992.

The Broadcasting Services Act (1992) (BSA) appears ill-equipped to deal with ‘new media’ services which challenge our understanding of what distinguishes a ‘broadcasting’ service from other services. The legislature’s move away from the original assumptions underlying the BSA (particularly the principle of technological neutrality) means that there are no unifying regulatory principles driving the ‘converging’ broadcasting and communications industries in Australia. The anticipated ‘policy and regulatory lag’ (Cutler, 1999) is becoming evident.

The first part of this paper uses examples of emerging new media services to illustrate how content which looks or sounds the same from the perspective of the audience may be regulated in different ways — depending on how that content is delivered. This will illustrate the mismatch between recent amendments to the BSA and the BSA’s ‘foundation principles’, particularly technological neutrality.

The second part of this paper will suggest some possibilities for the way ahead, particularly by reference to the proactive (and pro-convergence) proposals announced in the United Kingdom’s Communications White Paper (Department of Trade and Industry and Department of Culture, Media and Sport, 2000). The proposed UK scheme contemplates that regulation will be ‘rolled back’ when competition deems it unnecessary.

If we accept that ‘convergence involves a blurring of industry boundaries (for example, broadcasting and online services) and enables firms from traditionally separate industries (for example, broadcasting and telecommunications) to compete in new converged markets’ (Telstra, 1999), then as ‘new media’ content services start gaining a foothold in the Australian marketplace, we need to ask what Australia should be doing to address the ‘regulatory lag’.

Carolyn Lidgerwood is a Special Counsel at Gilbert and Tobin Lawyers in Sydney, where she specialises in broadcasting regulation.
The Divergence of the Regulatory Scheme

In the second reading speech for the *Broadcasting Services Bill 1992* (Senate Hansard) the then Minister for Transport and Communications said:

> We need new legislation capable of allowing the broadcasting industry to respond to both the complexities of the modern market place and the opportunities created by technological developments. Continuing to inhibit the natural development of this industry through outdated and cumbersome regulation will disadvantage consumers and be detrimental to the longer term prospects for Australia.

The limitations of the current Act have constrained development and structural development of the broadcasting industry, and thus diversity of choice for consumers.

The Minister also observed that some 20 substantial amendments had been made to the Broadcasting Act 1942 since 1983 and that further ‘ad hoc amendments will only add to its complexity and the potential for more loop holes to be created and exploited’. These comments explain why those who use broadcasting legislation on a regular basis may be occasionally struck by a feeling of *déjà vu*. Certainly in the last few years the scope of the BSA has expanded dramatically, with the enactment of literally hundreds of pages of amendments.

Since July 1998, the BSA has been expanded with the insertion of regulatory regimes for:

- regulation of online services — *Broadcasting Services Amendment (Online Services) Act 1999*;
- expenditure on new Australian and New Zealand drama programs by pay television licensees and channel providers — *Broadcasting Services Amendment Act (No. 3) 1999*;
- anti-hoarding (in relation to rights to events on the anti-siphoning list) — *Broadcasting Services Amendment Act (No. 1) 1999*;
- the licensing of digital datacasting services — *Broadcasting Services Amendment (Digital Television And Datacasting) Act 2000*; and
- international broadcasting services — *Broadcasting Services Amendment Act 2000*.

In addition, as detailed in the Liberal Party’s ‘Broadcasting for the 21st Century’ policy paper (Liberal Party, 2001), the Commonwealth Government proposes to amend the existing cross media and foreign ownership laws (at the time of writing, it is understood that draft legislation will soon be introduced).
These substantial amendments to the BSA have expanded the breadth of matters regulated by the Australian Broadcasting Authority (ABA) significantly. The ABA’s jurisdiction now reaches beyond the regulation of ‘traditional’ broadcasting services to the regulation of datacasting services provided in the broadcasting services bands and to online services hosted in Australia. However, these amendments leave the impression of having been progressively ‘cut and pasted’ into the BSA, without having been planned in an integrated way.

The consequence of these apparently *ad hoc* amendments has been a significant shift in the original assumptions underlying the BSA, namely that the regulation of broadcasting services would be technologically neutral, and that there would be a direct relationship between the degree of regulatory intervention imposed under the BSA and the ‘degree of influence’ which a type of broadcasting service was able to exert in shaping community views in Australia (see Explanatory Memorandum to *Broadcasting Services Bill* 1992).

The ‘degree of influence’ of the different categories of broadcasting service was assumed. For example commercial television services were deemed to be the most ‘influential’ category of broadcasting services in ‘shaping community views’ — presumably because of the type of content they provide, and their ubiquity. As the most influential kind of services, commercial radio and television broadcasting services are required to comply with content standards, content codes, wide ranging licence conditions, and are subject to stringent ownership and control rules. By contrast, those broadcasting services categorised as ‘open narrowcasting services’ are barely regulated at all, being subject to a class licensing scheme and minimum licence conditions, with no restrictions on ownership or control.

That the regulation of broadcasting services be ‘technology neutral’ was also a core objective in 1992. The definition of ‘broadcasting service’ in the BSA was originally intended to ensure that the method by which broadcasting services were delivered did not matter, with ‘broadcasting service’ defined as a ‘service that delivers television programs or radio programs …whether the delivery uses the radiofrequency spectrum, cable, optical fibre, satellite or any other means or combination of those means (BSA, subsection 6(1)). Before 1998, with the exception of the regime for satellite subscription broadcasting (BSA, Part 7), the principles of technological neutrality and degree of influence guided much of the legislative scheme for broadcasting and its implementation by the ABA.

With the recent amendments to the BSA, it can no longer be said that the driving principles of regulation under the BSA are technological neutrality or ‘degree of influence’.

As the Chairman of the ABA has observed, the current BSA ‘does not contain one system of regulation. In fact, it contains several widely divergent regulatory schemes’ (Flint, 2000). These so-called ‘divergent regulatory schemes’ for broadcasting, datacasting and online services also appear to be based on divergent regulatory policies. The question is whether this approach is sustainable, or whether the 10th year anniversary of the commencement of the BSA should coincide with a regulatory overhaul.
Departures from Technological Neutrality

The Explanatory Memorandum to the *Broadcasting Services Bill* 1992 explained that the new broadcasting legislation would provide ‘a simple regulatory regime for broadcasting services that applies irrespective of the technical means of delivery’ (Senate Hansard, at 3600). This is what is described in the broadcasting industry as the principle of ‘technology neutral regulation’.

As noted above, the Explanatory Memorandum explained that this principle supported the creation of ‘categories of service’ which would be described by their nature, rather than by their technical means of delivery. This would ‘ensure that the Act did not need constant amendment as technical conditions change’, as the BSA would allow for the ‘continuing development of the delivery of broadcasting services’ (Senate Hansard, at 3600).

The introduction of new categories of service for broadcasting was one of the key differences between the BSA and the previous regulatory regime (as contained in the *Broadcasting Act* 1942 and related legislation). It was intended that the ‘categories of service’ for broadcasting would determine the level of regulatory intervention which was appropriate for commercial broadcasting services, subscription broadcasting services, community broadcasting services, open narrowcasting services and subscription narrowcasting services. The regulation of the national broadcasting services (SBS and ABC) was ‘removed’ to the governing legislation of the national broadcasters, the *Australian Broadcasting Corporation Act* 1983 and the *Special Broadcasting Service Act* 1991 (Department of Transport and Communications, 1991), although the BSA does give the ABA a complaints handling function in respect of the national broadcasters.

The categorisation of broadcasting services was supported by the regulatory policy in section 4 of the BSA which provided that broadcasting services are to be regulated in a manner that, in the opinion of the ABA, ‘will readily accommodate technological change’ (BSA, paragraph 4(2)(b)). This regulatory principle has now been extended (by amending legislation) beyond broadcasting services to datacasting services, Internet content hosted in Australia and Internet carriage services supplied to end users in Australia (BSA, paragraph 4(3)(b)).

However, it can be argued that the application of the regulatory policy in section 4 of the BSA is quite meaningless in the context of recent legislative amendments that were specifically introduced to address technological change, particularly the use of digital technologies for datacasting in the broadcasting services bands (BSB). In such circumstances, the ABA has little flexibility to ‘accommodate technological change’ because the practical effect of the legislative scheme is not technologically neutral. This is explained in more detail below.

As observed in a submission to the Productivity Commission’s inquiry into broadcasting (Papandrea, 1999), landmark decisions in broadcasting since the BSA was introduced in 1992 have ignored the concept of technological neutrality. The first inroads into the principle of technological neutrality happened almost immediately after the commencement of the BSA, with the enactment of a ‘clumsy’ (Cutler, 1999) technology-specific regime for satellite subscription
broadcasting services (now Part 7 of the BSA, as amended by the Broadcasting Services (Subscription Television Broadcasting) Amendment Act 1992).

More recently, the undermining of the principle of technological neutrality has continued with the introduction of the regulatory regime for digital datacasting services. Under the present regime, a digital datacasting service which uses spectrum contained in the broadcasting services bands will be subject to the restrictive regulatory regime contained in the new Schedule 6 of the BSA. However, a service which looks like a digital datacasting service but is delivered outside the BSB may not be regulated under the BSA at all, despite the fact that it may look very similar from an audience perspective. This is because Schedule 6 of the BSA only applies to services provided in the BSB.

While it may not have been intended that the regulatory scheme for digital datacasting be ‘technology specific’ rather than ‘technology neutral’, this is the practical effect. When datacasting services begin to be provided in the BSB (whether by incumbent free to air broadcasters or by other parties who acquire datacasting transmitter licences at auction), they will be transmitted to domestic reception equipment from terrestrial transmission towers (that is, in the way that the free to air broadcasting services are usually provided). By contrast, datacasting services provided outside the BSB are likely to be delivered to set top reception equipment by cable or satellite or microwave (that is, in the way that subscription broadcasting services are provided).

If a datacasting-like service is provided by cable or by satellite using spectrum outside the BSB, it will not be regulated under the datacasting scheme in Schedule 6. Some examples of these kinds of services are illustrated below. By contrast, if the identical service is provided in the BSB, it will be subject to strict regulation.

Certainly, it is difficult to say that the service in the BSB will be more ‘influential’ than the non-broadcasting services bands service (that is, in the early years of free to air digital conversion, the proliferation of pay TV reception equipment is likely to make the non-broadcasting services bands service more ‘accessible’). This is not to suggest that non-BSB services should be regulated heavily — to the contrary. The point is that new media services which look the same from an audience perspective are regulated differently, and this is not justified by the principles of ‘degree of influence’ or ‘technological neutrality’.

If the policy reason for these regulatory distinctions is that datacasting services provided in the broadcasting services bands use scarce spectrum, then this would appear to be an issue for the allocation process or an issue to be covered through the imposition of licence fees — and not an issue which determines how the service should be regulated on an ongoing basis.

By moving the BSA away from technological neutrality, the ability of the regulator to respond to technological developments must be compromised. New content services are increasingly regulated according to how they will be or are delivered, not according to their ‘nature’ or ‘influence’ or how they appear to the audience. The question is whether this is an appropriate policy response to the regulation of converging services. Some examples for consideration follow.
**Example 1: Streaming audio and video**

The advent of ‘streaming video and audio’ services has meant that content previously identifiable as a ‘television program’ or a ‘radio program’ is now accessed from a PC (personal computer). However, streaming audio and video services are not broadcasting services if they are delivered using the Internet. This is because the Minister for Communications, Information Technology and the Arts (the Minister) has formally determined that ‘a service that makes available television programs or radio programs using the Internet, other than a service that delivers television programs or radio programs using the broadcasting services bands’ did not fall within the definition of ‘broadcasting service’ in subsection 6(1) of the BSA (Commonwealth Government Gazette, 2000).

This determination was made as part of a review that was required under the BSA. Section 216E of the BSA required the Minister to review whether, in the context of converging media technologies, streamed audio and video content obtainable on the Internet should be regulated as a broadcasting service. On one hand, making the determination was a way of reassuring providers of streamed audio and video services over the Internet that the ABA would not prosecute them for providing commercial broadcasting services without a licence. However, specifying ‘Internet’ delivery has the practical effect of drawing a technological distinction which inevitably raises further regulatory issues.

It is possible that if streamed audio and video services are accessed from ‘walled gardens’, they may not covered by the determination (and therefore, may be broadcasting services) whether they are delivered to a television set or to a PC (Nicholls, 2000)¹. The Minister’s determination applies only to programs ‘using the Internet’ and arguably, programs streamed over proprietary networks are not delivered using the Internet. The Internet does not discriminate on the basis of access (or ‘traffic’ flows), whereas walled gardens only allow traffic to flow within specified areas. In other words, walled gardens provide restricted ‘carriage’ access. It may only be a matter of time before this issue will need to be considered by the ABA. This is an example of how a move away from technological neutrality can lead to regulatory uncertainty. Clearly, it is not desirable for almost identical ‘new media’ services to be potentially subject to very different levels of regulation, but this is the kind of loophole which will always emerge when technological distinctions between different types of communications or content services are made by the legislative scheme.

**Example 2: Digital datacasting**

The introduction of digital broadcasting and datacasting and other interactive services will mean that services previously only available on computer screens

---

¹ A ‘walled garden’ is an environment that controls access to content and services on the World Wide Web by directing a user's navigation within particular areas (for more details, see http://whatis.techtarget.com)
will be available on television screens — with the use of set top boxes and infra-red keyboards. However, as introduced above, the regulatory regime in the BSA provides that some of these services (particularly datacasting) will be regulated only if they are delivered in the BSB. Otherwise, they may not be regulated at all.

For example, if a text-based service is provided outside the BSB, it won’t be a datacasting service, and it probably won’t be a broadcasting service either (text based services are excluded from the definition of broadcasting services in the BSA). Datacasting services do not fall into any of the broadcasting ‘categories of service’ because their determining characteristic will be as much about how they are delivered as how they look.

Subject to some exceptions, currently datacasting content licensees are prohibited from providing, in the BSB, content ‘genres’ including drama, sports, music, infotainment, documentary, comedy, quiz, children’s and light entertainment programs (see clauses 14 and 16 of Schedule 6 of the BSA). However, content such as Internet carriage services, interactive computer games and ‘ordinary’ e-mail is permitted and will be treated as a ‘datacasting service’ when it is provided by a datacasting licensee. Consequently, such content will be a datacasting service for the purposes of Schedule 6 of the BSA if it is delivered under the authority of a datacasting content licence, using the BSB. However, if it is delivered by use of spectrum outside the BSB, it will not be datacasting for the purposes of the BSA and probably won’t be a broadcasting service (depending on how the content is accessed).

If that is likely to be confusing, then it is just the beginning. Computer games consoles such as Playstation 2 will allow access to the Internet through television sets, personal digital assistants (for example, palm pilots) are able to download and display video (see Department of Trade and Industry (UK), 2000:7) and mobile phones will be able to be used to receive services which may sound very like commercial radio. These services won’t be regulated under the BSA, unlike identical content delivered by other methods.

The viewer, listener or user is unlikely to care much about this. Certainly, there are strong reasons why the delivery of content by ‘new media’ should not be regulated in the way that broadcasting services have traditionally been regulated. However, by not approaching the regulation of content services in a technologically neutral way, providers of content services may be prejudiced or favoured depending on what delivery mode they choose and whether or not they exploit a loophole in the regulatory scheme.

Under Part 4 of Schedule 6 of the BSA, datacasting licensees must develop and register codes of practice with the ABA, which will be periodically reviewed. The ABA may determine standards if it is satisfied that the relevant code of practice has failed. While codes and standards will not apply to an Internet carriage service or ordinary electronic mail (BSA, Schedule 6, clause 35), a licensee wishing to provide such content alongside other datacasting content, will also need to comply with separate code requirements (for example, Internet content hosts and service providers need to comply with the Internet Industry Association Codes registered with the ABA under Schedule 5 of the BSA).
If a person provides ‘datacasting-like’ content outside the BSB, then it will not fall within the regulatory scheme. However, the ‘broadcaster exemption’ from the definition of carriage service provider in section 93 of the *Telecommunications Act* 1997 may not apply (depending on whether the service can be described as a broadcasting service), which will mean that the service provider must participate in the Telecommunications Industry Ombudsman scheme, and will probably be deemed to be a carriage service provider. Other telecommunications industry codes of practice may also apply. The issue of the overlap of different regulatory regimes was raised in submissions to the Productivity Commission's broadcasting inquiry (for example, Austar, 1999). Under the current regulatory scheme, how content looks does not always determine how it is regulated.

The application of different regulatory regimes to new services which look the same from an audience perspective does not seem to be an appropriate approach for a converging communications environment. Certainly, ‘regulatory risks’ may arise if regulations in converging sectors are incompatible (Productivity Commission, 2001). In that context, the key question would appear to be when and how the regulation of converging services can be made more consistent.

At the time of writing, the operation of Schedule 6 of the BSA (containing the datacasting services regulatory scheme) is under review. An Issues Paper was released by the Department of Communications, Information Technology and the Arts (2001). The review is limited to the regulation of datacasting services provided in the BSB, and does not address the broader issues set out above.

*Not influential, just unfamiliar*

The Explanatory Memorandum to the Broadcasting Services Bill 1992 explained that the ‘degree of influence’ test enunciated ‘the underlying philosophy to be pursued in the administration of the Act’:

The level of regulation to be applied to any particular type of broadcasting service is to be determined according to the degree of influence that that type is able to exert in shaping community views in Australia. Thus a high level of regulation is to apply to commercial broadcasting services as those services are considered to exert a strong influence in shaping views in Australia. At the other end of the scale, narrowcast broadcasting services are expected to play a minor role in shaping views in Australia and will be subject to low barriers to entry ….

Since that time, there has been a practical shift away from the principle of regulating services according to their degree of influence. This may be an unintended consequence of recent amendments, given that the regulatory policy statement in section 4 of the BSA was amended so that the ‘degree of influence’ test is used to justify the regulatory intervention in relation to online services and datacasting services. Subsection 4(1) of the BSA, as amended, now states:
The Parliament intends that different levels of regulatory control be applied across the range of broadcasting services, datacasting services and Internet services according to the degree of influence that different types of broadcasting services, datacasting services and Internet services are able to exert in shaping community views in Australia.

Despite this statement, while the ‘degree of influence’ test still applies to some extent in shaping traditional broadcasting regulation, it appears to have lost its central importance in the context of non-broadcasting services which have been brought within the scope of the BSA. For example, it would not appear that the test justifies the current regulatory regime for datacasting (which is currently under review, as noted above), as the legislative scheme effectively strips datacasting services of any influence they may have had in the absence of the ‘genre’ rules. It may be that the amendments to section 4 of the BSA were more ad hoc than representative of the true nature of the new regulatory scheme.

Similarly, the ‘degree of influence’ test is unlikely to justify the legislative regime for online services. In the course of the Productivity Commission’s Broadcasting Inquiry, the Communications Law Centre (2000) suggested that the test was particularly undermined by the online amendments, ‘whose interventions are informed less by any understanding of the degree of influence of on-line media than by the kind of naively apocalyptic vision of media power and influence which so often accompanies the introduction of new technologies’. It could also be suggested that the complicated scheme for expenditure on new Australian or New Zealand drama by pay TV operators and channel providers is also disproportionate to the amount of influence which those services individually wield, especially in light of the administrative burdens imposed by the ABA’s reporting requirements.

The degree of influence principle sounds rather hollow when these practical examples are considered.

If what determines the level of regulation upon a new class of content service is the fact that service uses scarce spectrum (that is, in the BSB) then shouldn’t that be stated as the determining regulatory principle? If the real driving factor of regulation of on-line services is the protection of children, shouldn’t that principle replace the degree of influence principle? If the imposition of the expenditure scheme on the pay TV industry is an industry protection measure for the local production industry, then call it that — not regulation which is justified because of the supposed degree of influence that pay TV drama services wield. In that context, the issue is not whether the central regulatory principles underpinning the BSA need to be restated and simplified, but when this should happen.

**A Suggestion for the Way Ahead**

This year marks the tenth anniversary of the commencement of the BSA. Given this, and the fact that ‘converging services’ are obtaining a foothold in the Australian media landscape, it is appropriate to consider whether the existing regulatory approach in the BSA is sustainable.
Under the digital television legislation, a number of reviews were required to be conducted prior to 1 January 2001. Clause 59 of Schedule 4 of the BSA required the Minister to cause a review to be conducted of whether ‘any amendments of laws of the Commonwealth should be made in order to deal with convergence between broadcasting services and other services’.

The Report of the Convergence Review (Department of Communications, Information Technology and the Arts, 2000) favoured the retention of industry-specific regulation. In relation to broadcasting, the Report concluded that existing structures were likely to persist until ‘digital technology substantially penetrates the broadcasting industries’, and that ‘this was several years off’. As a result, the review concluded that ‘the structure of the broadcasting legislation will remain sound for some time’ (Department of Communications, Information Technology and the Arts, 2000:8).

The approach of the Convergence Review was that the development of converged regulatory structures should be ‘evolutionary’. However, this is a less than proactive approach, in comparison with other jurisdictions (and indeed, with other industry sectors, as discussed below).

What the Convergence Review failed to do was to produce a vision for the ‘convergence’ of broadcasting and communications regulation. By comparison, the UK Departments of Trade and Industry and of Culture, Media and Sport have jointly published a White Paper titled ‘A New Future for Communications’ (December 2000), which sets out the UK Government’s response to regulation in a converging environment.

At the centre of the White Paper’s recommendations is the proposal to create a high level set of principles and objectives for the regulation of content across all electronic communications to be administered by one merged regulator, an Office of Communications (OFCOM) that will deal with competition, carriage and content issues. The White Paper states:

Taking full account of the differences between services and people’s expectations of them, OFCOM will be responsible for maintaining content standards in the electronic media. It will develop Codes underpinned by statute for the most pervasive broadcast services, and work with industry to ensure effective co- and self- regulatory approaches to protection for other services, such as the Internet, where they are appropriate.

This framework can be compared with a simplified and updated version of the scheme currently contained in the BSA.

The White Paper proposes (8.5.1) that OFCOM will have the following central regulatory objectives:

- protecting the interests of consumers in terms of choice, price, quality of service and value for money, in particular through promoting open and competitive markets;
• maintaining high quality of content, a wide range of programming, and plurality of public expression; and
• protecting the interests of citizens by maintaining accepted community standards in content, balancing freedom of speech against the need to protect against potentially offensive or harmful material, and ensuring appropriate protection of fairness and privacy.

The White Paper states (8.5.2) that OFCOM is also to give weight in all its activities to the promotion of efficiency, including efficient use of spectrum and telephone numbers, and the promotion of innovation. Effective regulation is seen as the main challenge in a converging communications environment. The White Paper proposes that OFCOM will develop and maintain the necessary rules in full consultation with industry and consumer representatives, within this broad framework of guiding principles (established in statute).

Importantly, it is proposed that OFCOM will have a duty to keep industry sectors under review and to ‘roll back regulation promptly where increasing competition renders it unnecessary’ and to ‘encourage co-regulation and self-regulation will best achieve the regulatory objectives’ (8.11). The White Paper states (at 8.11.2) that OFCOM will give full weight to the principles of proportionate regulation. This approach should be contrasted with the conclusions in the Australian Convergence Review (Department of Communications, Information Technology and the Arts, 2000).

It is worth noting that the House of Commons Select Committee on Culture, Media and Sport (2001) considered the White Paper in the context of the dramatic changes flowing from the introduction of digital television, and the growth of the Internet and mobile telephony in the UK. The Select Committee observed that:

The functional divisions of current regulators are ceasing to reflect the realities created by technology. Current operators, and in particular those concerned with networks and with innovation, face a confusing, overlapping and inefficient regulatory environment. In these circumstances, the case for a new regulatory framework of the kind proposed in the White Paper is overwhelming (Part II:13).

It can be argued that the same observations apply to the Australian communications environment.

What is most appealing about the UK proposals is the fact that the regulatory system aims to ‘provide industry and consumers with certainty, while being flexible enough to recognise the differences between different services and to respond to rapid changes in technologies, services or public expectations’ (White Paper:6.3.8). The proposed model recognises that different levels of regulation are appropriate for different kinds of services, but that these levels should not be set in concrete.
Draft legislation implementing the White Paper proposals is currently being prepared, although it is not expected to be published until early 2002 (Wintour, 2001).

**Conclusion**

As noted above, the White Paper recommended the creation of a high level set of principles and objectives for the regulation of content across all electronic communications. It is worth noting that (by contrast), the BSA contains not only the regulatory principles in section 4 (referred to in this paper), but 18 ‘objects of the Act’ in section 3, as well as detailed policy objectives for digital television conversion in Schedule 4.

In that context, the simplicity of the proposed UK approach is appealing — that is, establish some over-riding principles, and then trace out appropriate objectives in the relevant codes and conditions for the most ‘pervasive’ services. While this approach (if applied in Australia) may result in less regulation for those kinds of services which are currently deemed to be the most ‘influential’ (for example, commercial radio), and some code-based regulation for services that may not currently be regulated at all under broadcasting legislation (for example, computer games services), this approach would move beyond the historical distinctions which characterise much of the present regulatory scheme.

Recently, it was observed that broadcasting regulation in Australia reflected a legacy of *quid pro quos*, a history of political, technical, industrial, economic and social compromises (Productivity Commission, 2000). Broadcasting regulation has never been about a reliance on market forces or the promotion of competition. A stark comparison can be drawn between broadcasting regulation and telecommunications regulation in Australia, for example. The telecommunications sector has been steadily deregulated since the late 1980s, and in 1997 barriers to entry were significantly reduced. As the Productivity Commission (2001) observed:

> The convergence of broadcasting and telecommunications accentuates the paradoxically pro-competitive orientation of policy towards traditional telecommunications and the protective pall of regulation that surrounds broadcasting.

It is inevitable that the boundaries between the broadcasting and telecommunications industries will become increasingly blurred in the short term. As the platforms available for ‘convergent’ services like video on demand services, messaging services and Internet services become more plentiful (for example, as a result of technological improvements), the separation of the broadcasting and telecommunications regulatory regimes is likely to be more difficult to sustain.

Even recognising the historical differences between the regulatory schemes of the UK and Australia, the UK proposals are worth serious consideration. Given
that the ownership and control rules in Australia are currently under review, it
would be a lost opportunity if this process did not form the impetus for a broader
review of the broadcasting regulatory regime. The UK has approached the task in
the reverse order — with a review of the UK’s complex media ownership laws
currently underway (Harding, 2001).

A proactive convergence policy framework must be the way ahead, rather
than ‘reactive’ legislative responses to broadcasting regulation which have been
implemented in recent years in Australia. As the General Manager of the ABA
has noted (Tanner, 1998) ‘in a converged world, the prognosis for much
traditional broadcasting regulation looks grim’. We can conclude with the
observation that in a converged world, the prognosis for the principles of
proportionate regulation should be bright.

References

Austar United Communications Limited (1999), ‘Submission to the Productivity Commission

Australian Broadcasting Corporation Act 1983.


Broadcasting Services Amendment Act (No. 1) 1999.

Broadcasting Services Amendment Act (No. 3) 1999.

Broadcasting Services Amendment (Digital Television And Datacasting) Act 2000.

Broadcasting Services Amendment (Online Services) Act 1999.

Broadcasting Services Amendment Act 2000.


Commonwealth Government Gazette (2000), Government Notices, 27 September (See also the

Communications Law Centre (2000) ‘Submission to the Productivity Commission Review of the
Broadcasting Services Act 1992 and associated Licence Fees Acts’, June,

Copyright Amendment (Digital Agenda) Act 2000.

Cutler, T. (1999) ‘Avoiding conflict between communication policy and convergence’, Agenda,

Department of Communications, Information Technology and the Arts (2000), Convergence Review,

Department of Communications, Information Technology and the Arts (2001), ‘Review of the
Operation of Schedule 6 of the Broadcasting Services Act 1992 (Datacasting Services)’ Issues Paper,


Senate Hansard (1992), Second Reading Speech to the Broadcasting Services Bill 1992, 4 June at 3599 (Senator Collins).

Special Broadcasting Service Act 1991.


All views expressed in this paper are personal. An earlier version of this paper was presented to the ABA’s Radio, Television and the New Media Conference (Canberra, 3-4 May, 2001). The author gratefully acknowledges helpful comments by two anonymous referees.
National Competition Policy: Some Issues

Fred Argy

Introduced as part of the Hilmer reforms of 1995, national competition policy (NCP) is a cooperative Federal-State attempt to broaden the scope of competition and develop a more coordinated approach to reform. The main provisions are set out in the Competition Principles Agreement (CPA).

Some competition reforms were already in place before NCP was introduced. For example a few individual governments had earlier opened up infrastructure monopolies such as gas, electricity, water and transport to wider competition. The NCP framework took these reforms under its wing and embraced many new initiatives which sought to:

- extend the Trade Practices laws to certain government-owned and private businesses previously exempt;
- promote competitive neutrality between private and government-owned businesses;
- encourage governments to undertake systematic reviews of all anti-competitive elements in existing legislation; and
- ensure reasonable access by competing businesses to core monopoly controlled infrastructure such as electricity cables, railway lines, gas pipelines and airports.

The broad principle underlying NCP is that restrictions on competition should be removed unless they can be shown to be in the public interest. The National Competition Council (NCC) has been given the role of advising the Federal Treasurer on whether the aims of the reform program are being met and if not, whether special Commonwealth revenue grants (‘competition’ or ‘bonus’ payments) should be withheld.

Concerns have been expressed about the competition reform program on three main grounds:

- it does not allow state authorities adequate discretion in relation to policy development, review and pace of implementation;
- its effective contribution to economic efficiency is small; and
- it gives too much weight to efficiency relative to other societal goals.

The main aim of this paper is to examine critically these concerns.

Fred Argy is a former government policy adviser and currently a Visiting Fellow at the ANU and Adjunct Professor, University of Queensland.
Democratic Legitimacy of NCP Processes

Some critics (for example Quiggin, 1998:16-17) say that NCP allows the state governments too little discretion in pursuing their own social priorities and that processes of consultation and implementation are inadequate. They therefore question its democratic legitimacy. There are several strands to this concern.

First, some critics allege that only limited consultation occurred when the policies were first formulated. This claim is not explored further in this paper. Second, it is argued that NCP arrangements do not allow individual governments enough political flexibility to trade-off some efficiency for other values in response to community preferences. This is an important issue; it is discussed later in the paper. A third strand is that non-elected officials and advisers have excessive influence relative to state politicians. This concern relates to the powerful role played by the NCC in monitoring the reforms and advising the Federal Treasurer. Finally it is argued that states rights are being infringed by giving the Commonwealth Government powers to impose financial penalties on the states when it is dissatisfied with the results.

The need for State Governments to have a bigger say in the policy development process was recognised by federal and state ministers at their November 2000 meeting. They agreed on a number of changes designed to enhance the opportunity for states to make their case where the NCC recommends a penalty and before the final decision is made on competition payments (Banks, 2001:13). State authorities are still required to document their public interest reasons and to ensure that the outcome is within a range which ‘a properly constituted review’ might consider ‘reasonable’ on the information available to it. But they are being allowed more discretion in determining whether existing practices are in the public interest.

To go further and totally abolish the right of the Federal Government to withhold competition payments — remove altogether the element of incentive (as proposed by some, for example Quiggin, 1998) — would leave NCP without a framework to ensure that the rules of the game are observed. It would weaken confidence in the transparency, independence and objectivity, and hence the consistency and integrity, of public interest assessments.

On the other hand, it is equally untenable to argue that the role of the states needs to be substantially reduced in order to speed up the process of reform (Harman and Harman, 1996:2). Such critics overlook the extreme difficulties of winning political cooperation and support without allowing states discretion on how best to achieve agreed outcomes (Keating and Wanna, 2000:139).

NCP and Economic Efficiency

Some economists believe that reforms like NCP make little or no contribution to economic efficiency. This criticism is discussed here firstly at the theoretical level and then by specific reference to NCP.
Theory

A reform improves economic efficiency if winners are capable of compensating the losers and still remain better off. (A more restrictive concept of efficiency requires that winners actually compensate losers but policy-making would be paralysed if it always had to strive for win/win outcomes.). A recent survey of economists found an overwhelming majority was of the view that pro-competition policies will tend to improve economic efficiency (Argy, 2001). Such consensus is rare in economics. It stems from a belief that competitive markets give consumers wider choice and lower prices and give sellers stronger incentives to:

- minimise their costs and cut out waste (thus making less calls on national productive resources to produce any given output);
- innovate and adapt quickly to changing circumstances;
- pass on cost reductions to consumers and target their specific preferences; and
- channel goods and services to those consumers who value them most highly.

However economists are also aware of three ‘traps’ in the theory. First, free markets do not always work perfectly. There may be information asymmetries, high transaction costs involved in collecting and evaluating information, learning curves, herd mentality affecting buyers and sellers, and so on. More importantly from the viewpoint of this paper, the market may not be able to sustain several producers if there are potentially large economies of scale and natural barriers to entry. In such circumstances competitive markets may not be viable in the long term and can lead to a concentration of private market power. This would in turn require extensive regulation and would not necessarily produce good efficiency outcomes.

Second, unleashing the forces of competition across the economy may entail economic adjustment costs in the short term and these costs need to be set against long term efficiency gains. For example, many unskilled workers displaced by structural change and reform remain unemployed for fairly long periods and suffer severe personal costs (Borland and McDonald, 2000).

Third, competitive markets may not produce efficient outcomes in the long term if there are ‘negative externalities’, that is, adverse economic welfare consequences not fully accounted for in the price and market system. Such externalities can occur, for example, if the competitive process leads to erosion of a country’s natural capital or increased traffic congestion and airport noise or increased job insecurity and stress.

Even the most enthusiastic supporters of competition policy acknowledge that if market imperfections, adjustment costs and negative externalities are large enough and are ignored, they can conceivably leave little or nothing by way of net efficiency gain.
Potential efficiency gains from NCP

NCP is essentially a mechanism for freeing up markets and facilitating greater competition. How well does it deal with the three ‘traps’ outlined above?

Some public service or utility markets targeted by NCP operate in imperfect markets. They may not lend themselves to additional competition, either because of economies of scale (as with core networks) or because of the high costs of collecting and evaluating information (as with health and education). But such imperfections are only found in a few of the markets subject to NCP and can be separately accommodated, for example, through price and access regulation (Fels, 2001), without abandoning the whole reform program.

Implementation of NCP is also likely to involve significant adjustment costs on the way. For example, the Productivity Commission (1999:250) acknowledges that, although in the long run there should be positive employment effects through price reductions and real income increases, ‘the early direct effects of NCP reforms on employment have been adverse’. Victoria’s Latrobe Valley suffered sustained job losses as a result of electricity reform, although employment levels in the Valley region are now tending to recover significantly.

While short term employment disruption is unavoidable as NCP begins to bite, good governance can ensure that it does not lead to long term, structural unemployment — an issue the paper returns to later.

There are also transitional costs stemming from water reform, for example, the lack of success in addressing water property rights is eroding security for farmers in financial negotiations. Again, such costs can be minimised with appropriate policy action. Similarly, any negative externalities generated by NCP can in general be prevented or offset without destroying the basic efficiency rationale of the reform program.

For example, Hamilton and Denniss (2000) warn that the application of competition policy to the electricity sector could, by encouraging more electricity production, add substantially to the combustion of carbon-based fossil fuels and greenhouse gas emissions. However, the problem here lies not with competition policy per se but with environment, energy and utility pricing policies. For example, the price of electricity could incorporate a special carbon charge to ensure it is not below the socially optimal (and consumption is not above optimal). Or the tax system could be restructured so as to encourage a shift from coal-fired to gas-based fuels and so reduce the ‘carbon dioxide intensity’ of electricity generation. More generally, a balanced national energy policy could be implemented in conjunction with NCP. A review of the energy market foreshadowed by the Council of Australian Governments (COAG) is to include greenhouse emissions in its terms of reference. Limited steps are also being taken to facilitate the trading of emission credits.

In any case the environmental externalities generated by competition policy are not all negative. For example, cost minimisation in electricity reduces demand for natural resources per unit of output. Similarly, water reforms, such as the application of fuller cost pricing to water use, the allocation of clear rights to use
of water and the introduction of tradeable allocations of water rights, are likely to prove favourable for the natural environment, and especially our river systems. Again, the wider choices offered by NCP should make it easier for buyers to purchase ‘green’ energy from renewable sources like natural gas.

On a different plane, it has been argued that reforms like NCP which tend to intensify the competitive environment can put pressure on firms not to exceed legally required safety standards and working conditions and to demand more of their workers. The market liberalisation reforms of the 80's and 90's have been associated with a rise in casuals, more widespread job insecurity and increased work intensity and stress, as well as higher structural unemployment. The evidence of deterioration in workers' quality of life is now quite extensive (for example, see ACIRRT 2001; Pocock 2001; Kelley Evans and Dawkins, 1998), although it is still subject to debate (Murthough and Waite, 2000; Wooden, 1999). Being based on opinion surveys, the evidence is perceptual in character. But perceptions are important in such cases.

The potential efficiency costs of such effects on workers are far from insignificant. For example, when asked, typical low-income workers say that they would be happy to earn one third less to move from an insecure to a secure job (Kelley and Evans, forthcoming). If governments are concerned (as they should be) about what is happening to the quality of life of workers, it does not mean they have to walk away from competition reform. They can continue to implement reform and then neutralise some of its undesirable effects by other more direct means such as establishing minimum economy-wide workplace standards.

In short, market imperfections and failure do not per se destroy the efficiency rationale of competition policy. However they have a message for policy-makers. The efficiency benefits of competition reforms like NCP cannot be taken for granted; they must be assessed on a case by case basis and the reforms often need to be supplemented by other policies. Sometimes the externalities are so complex and difficult to anticipate and control that they are best handled under a regulated public monopoly (Argy et al., 1999; King and Pitchford, 1998).

Actual NCP efficiency returns

It is difficult to evaluate the actual impact of NCP on efficiency and costs. For one thing, implementation has progressed more slowly than expected. Rail reform is lagging and there has been some slippage in achieving a national gas and electricity market and in reviewing anti-competitive practices in the previously exempt private sector. The slow pace of reform reflects the sheer technical complexity and ambitious scope of the exercise, but it also stems from community and political concerns about the pace of change. Whatever the causes, NCP remains an unfinished project.

Many analysts point to big reductions in prices of gas and electricity for large users (Willett, 2001) and to other benefits such as lower conveyance fees, savings in water usage and reduced duplication (NCP, 1999:9). Some also argue that for most electricity customers the incidence of supply interruption has been reduced.
However, some of the initial efficiency gains from NCP may not be sustained. Initially, electricity prices fell but this was partly because the industry was deregulated at a time of excess capacity. With this excess capacity eroding, and little or no generating reserve capacity to draw on, the expectation now is that retail electricity prices will rise appreciably in the near future — at least until new capacity comes on stream. Although the recent world wide downturn in demand for energy and falling oil prices may be fortuitous in easing the transition, the electricity reforms are about to be tested to the full.

Longer term, there are doubts about the competitive structure of the industry. ‘High regional pool prices could indicate that the generation market is too thin and that individual generators have substantial market power … the market is not currently working as well as it should’ (Willett, 2001). The NCC attributes this to inadequate inter-regional competition, a cumbersome regulatory framework and inadequate competition between generators. The challenge for regulators will be to reduce monopoly power and excess profits and yet ensure adequate incentives for private electricity operators to invest in new transmission links. Fortunately, Australia does not face the same fundamental problems as in California, where wholesale prices were freed while retail prices were frozen. Australian authorities are aware of the need for regulated prices to be regularly and flexibly adjusted.

Clearly, the original form envisaged for the national electricity market will require some modification. In the meantime, Queensland (unlike other states) has decided not to proceed with retail contestability.

Despite all these transitional and structural difficulties, most economists are confident that NCP will appreciably improve efficiency in the long term. Various independent reports (cited by Willett, 2001; and Banks 2001:3-4) have estimated that the electricity reforms will improve GDP by several billions of dollars. These projections are consistent with Australia’s own past experience with microeconomic reform. In the decade of the 1990’s, Australia’s multi-factor productivity growth was double the rate in the preceding decade and a half and 0.8 percentage points higher than the OECD average, whereas it was below or equal to OECD in the two previous decades. These productivity gains partly reflect cyclical recovery effects that some analysts believe to be dominant (Quiggin, 2001). However most economists see evidence of a decisive improvement in productivity trend (Dowrick, 2001) and, while they attribute some of the gains to improvements in technology and human capital and better macroeconomic management, they assign a major role to the microeconomic reforms of the 80’s and 90’s. Many of these reforms (for example, tariff reductions, deregulation of financial services, transport and communications and contracting out in the public sector) had as their principal intent to increase competition.

Model-based estimates are fraught with uncertainty and the precise numbers they churn up cannot be taken seriously; but the broad impressions they convey are in line with what one might theoretically expect. Apart from widening consumer choice and giving sellers the right sorts of incentives, NCP is likely to have a positive disciplining effect on state governments by instilling in them a culture of rigorous justification of the need for and design of new business
regulation’ (NCC, 2001:Foreword). In short, while the results so far are patchy and uncertain, NCP can reasonably be expected to yield important efficiency benefits in the longer term.

Consistency of NCP rules

If efficiency is the main aim of NCP, then it needs to be applied consistently. Anti-competitive practices by, say, pharmacists, newsagents and some professionals have not been subjected to the same rigorous cost-benefit evaluation on public interest criteria as other restrictive practices. Political sheltering erodes the integrity and even-handedness of the system.

Equity as an Important Dimension of Community Well Being

A gain in economic efficiency is basically value-free. It simply widens the choices available to the community. The community can use the benefits to increase consumption of goods and services or for other purposes such as extend leisure, improve the quality of life and restore the environment. Whatever a nation’s values and goals, these can be achieved better if national resources are used and allocated efficiently. So if NCP enhances efficiency it also has the potential to increase community well-being (social welfare). However realisation of this potential will depend on how the benefits and costs are distributed.

Although economists are taught not to make interpersonal utility comparisons, value judgments in policy formulation are inevitable. The so called ‘neutral’ assumption that everyone has the same marginal utility (which is the implicit premise when a reform is assessed exclusively in terms of aggregate GDP outcomes) is itself fraught with value judgments.

If value judgments are inevitable, governments should try to reflect predominant values as far as possible. These values are not easy to determine but opinion surveys point to three pertinent conclusions. First, most Australians seem prepared to sacrifice some economic growth in order to achieve less inequality (Kelly, 2000:230). Second, Australians seem prepared to pay higher taxes in exchange for better government services that promote equal opportunity (Withers and Edwards, 2001; Kelley and Evans, 2001). Third, a person's perceived well being seems to depend at least as much on changes in relative income and quality of life as on absolute changes (see Oswald, 1997; Frank, 1997).

Prima facie this suggests that, to ensure efficiency benefits are translated into community wellbeing, low income people should ideally share (at least in part) in the incremental benefits of reform — and not just over the ‘long term’.

It is unlikely that NCP has had much impact on distribution so far. But it does have the potential to do so increasingly in the future. Some of the effects will be socially progressive. Increased competition in the utilities and professions should tend to reduce incomes inequality by breaking down market power, forcing prices to fall in line with productivity improvements and eroding privileged positions. A study by the Productivity Commission has found that ‘productivity
gains at the industry level have predominantly been passed on in the form of lower
prices. This is particularly true of the 1990's, suggesting that stronger competitive
pressures have been at work' (Parham et al., 2000:xiv). The widespread use of
utilities like electricity and gas means that lower prices are of special value to
poorer households.

However there is a risk that, in the absence of active policy intervention, the
regressive effects will outweigh the progressive effects. A more competitive
environment almost by definition favours strong, well endowed, adaptable and
competitive people and hampers those most vulnerable to change. The evidence
(Katz, 1998:33-8; OECD, 2001:Table 2-1) appears to bear this out — across
countries, there was a tendency for earnings inequality (before taxes and transfers)
to widen in the late 80's and 90's in response to market liberalisation reforms.

In Australia too the increased pace of change, deregulation and liberalisation
in the 80's and 90's appear to have led to an increase in earnings dispersion, with
lower paid workers falling behind (Parham et al., 2000:135-6). But Australian
governments recognised this danger and took measures through social security
payments, the progressive tax system and ‘social wage’ (that is, non-cash benefits)
to offset the increase in market inequality (Harding, 2001; Saunders, 2001). As a
result, net incomes inequality, conventionally measured, did not significantly
increase over the last two decades. But without on-going government intervention
(both passive and social assistance) inequality would certainly have increased, as
it did in many other reformist countries such as the UK, NZ and the US.

Similarly, competition policy can increase regional inequality. A more
competitive environment creates incentives to centralise costs and roll back
uneconomic services, thus affecting the smaller regions. A study covering the
period 1986 and 1996 found ‘dramatic increases in regional inequality’ especially
between states and post codes within states (NATSEM, 2000:2). A more recent
study suggests that, at a national level, regional inequality has stopped increasing
in recent years, probably reflecting falling unemployment and rising farm incomes
(Harding, 2001). The Productivity Commission (1999:257) believes that most
statistical regional divisions will be positive winners from the major NCP reforms
relating to utilities (also see Banks, 2001:6), but there are likely to be transitional
and even sustained adverse effects on smaller specialised geographic regions.

If governments believed that a particular reform would hurt particular
sections of the community and wanted to prevent or temper this effect, they could
take offsetting action either ex ante or ex post. In the case of NCP, one way —
but as we will argue not necessarily the best way — would be to use the Public
Interest mechanism to justify retention of the restrictions on competition.

**Deficiencies in the Public Interest Test**

The Public Interest Test (PIT) is set out in clause 1(3) of the Competition
Principles Agreement. It requires governments to take into account not only the
effects of NCP on economic efficiency (such as competitiveness of Australian
industry, efficient allocation of resources and economic development) but also
ecological sustainability;
• social welfare and equity, including community service obligations;
• compliance with prescribed standards on occupational health and safety, wages and working conditions;
• reasonable equality of access to essential services; and
• regional development.

Nor is this clause meant to be exclusive. It is open to governments to take account of matters not specifically listed in the clause ‘such as the impact on specific communities, including adjustment costs’ (Samuel, 2001:4). As well, critics of NCP are frequently reminded that the Agreement, although it does require competitive neutrality, does not insist on privatisation of government assets, nor does it insist on compulsory competitive tendering and contracting out.

The NCC therefore believes the PIT offers an adequate safeguard against over-preoccupation with efficiency relative to other societal goals. It argues that ‘the NCP agreements give social and environmental values no less weight than financial considerations in determining where the public interest lies. … All public interest considerations intrinsically carry equal weight’ (NCC, 1999:20). But critics are not satisfied. Quiggin (1998:8) argues that ‘NCP differs from earlier attempts to promote competition by virtue of the assumption that competition is always and everywhere desirable and that where competition is in conflict with other values, there should be a presumption in favour of competition’.

This may over-state the weight given by NCP to competition but it raises a legitimate concern. Despite the broad ranging character of the PIT and the large discretion on social policy allowed individual governments, there are doubts about the effectiveness of the Test in establishing an acceptable balance between efficiency and equity. These doubts arise from two features of the PIT.

First, the onus lies with opponents of reform to prove, through rigorous and independent public interest reviews, that existing restrictions on competition are in the public interest. The NCP framework presumes that competition serves the public interest unless it can be shown otherwise — that is, that ‘competitive outcomes deliver greater benefits than non-competitive outcomes in the absence of evidence to the contrary’ (Samuel, 2001:3). This approach is not the usual practice with economic reform. Normally it is up to reform proponents (including the Productivity Commission) to show that a particular change will be worthwhile. By assuming that existing restrictions on competition are against the public interest unless shown not to be, NCP is an exception to this rule. True, the Trade Practices Act uses an approach similar to the NCP’s to deal with anti-competitive conduct. But in that case the aim is to curb the abuse of private monopoly power for private purposes — whereas in the NCP’s case it is to curb the potential for abuse of political power. The two cases are qualitatively different.

The Chairman of the Productivity Commission, Gary Banks, has noted another feature of the PIT which may compound the bias in favour of competition and efficiency. In a recent talk, he warned of the ‘danger that only the measurable (what can be easily quantified or valued) will be influential in decision-making’.
Since effects on distribution, the environment or quality of life are not easily measured, quantified and demonstrated through ‘rigorous, independent reviews’ (as required by the Agreement), such values will tend to start with a disadvantage relative to productive efficiency (usually measured by GDP per hour or per head). More generally, Banks disputes the view that social and environmental considerations carry ‘equal weight’ with other public interest criteria. ‘This could be misconstrued as them having equal importance in all cases. It may be better to describe the criteria as having equal status’ (Banks, 2001:9).

The effect of these features of the PIT is to give proponents of competition reform an advantage over opponents. If this is a concern to governments, there are a number of possible policy responses. They are discussed in the next section.

**Possible Policy Responses**

The President of the NCC, Graeme Samuel (2001:8), accepts that governments have a responsibility to ‘ensure the benefits (of competition policy) are shared equitably’. He also believes that ‘to date governments have responded poorly to this responsibility’. There are three ways to more effectively meet this responsibility.

*The ‘reform dilution’ approach*

The first would be to try to amend the Public Interest Test provided in the Agreement so as to make it easier for governments to show that existing restrictions on competition are in the public interest. This can be called ‘reform dilution’ because it involves deliberately foregoing some efficiency benefits in order to safeguard other goals (see Argy, 1999).

There are various ways the PIT can be amended to achieve the desired result. Quiggin (1998:17-19) suggests restructuring of the onus of proof to take ‘explicit account of all the criteria set out in the CPA s.1(3) as well as any other relevant factors, including losses to employees through reductions in wages and increases in work intensity’. Less ambitiously, the Federal Government has foreshadowed some tinkering with the PIT. The Deputy Prime Minister is reported to have said that ‘the public interest test needs to take proper account of the needs of rural communities and the costs of reform for regional communities’ (Lewis, 2001).

The option of diluting the PIT, while technically viable, would add a lot of complexity to the process. More importantly, it would almost certainly involve a significant sacrifice of efficiency if it made it too easy for governments to retain market-distorting devices such as restrictions on competition to advance their social goals — instead of looking for more cost-effective devices.

*The ‘exemption’ approach*

A second possible approach, potentially involving an even larger sacrifice of efficiency if carried too far, would be to legislatively exempt a wide range of activities from the provisions of NCP (as is already done for most pharmacists).
Such an approach would side-step the need for rigorous, objective and transparent cost-benefit analysis. There would therefore be a risk that reform would be suppressed merely to satisfy vocal vested interests. Moreover, like the reform dilution approach, it would involve employing a market-distorting instrument.

That said, there may be a case for excluding from NCP such areas as the provision of health and education. Apart from being socially sensitive, these are sectors where competitive markets do not always work well — for example, where consumers are not adequately informed and there are high transaction costs involved in acquiring the necessary information.

The ‘reform with smoothing’ approach

A third possible response would be to leave the PIT as it now stands and address the equity and environmental concerns associated with competition policy through direct budgetary means. This would mean providing special assistance in addition to generally available assistance measures (such as those administered by Centrelink). The additional assistance need not be of the income support variety, nor need it be solely ‘compensatory’ in its intent. It can include what the OECD has called ‘active social policies’ which seek to encourage a creative change in the behaviour of assistance recipients (for example, inducing them to move to where the jobs are or to enhance their skills or gain some work experience). Such active policies may include for example: adjustment assistance, expenditure on regional infrastructure, equal opportunity measures and active labour market programs (such as wage subsidies, training, improvements in job placement machinery and intervention to promote welfare to work). Active policies are socially more rewarding in the long term because they get at the root cause — market inequality — and eventually reduce the need for passive welfare. They are also economically more efficient (OECD, 2001; Forsyth, 1999; Argy, 1999).

Such an approach to economic reform can be called ‘reform with social smoothing’ because it seeks to safeguard wider community values through the most efficient and effective means available, while maintaining the momentum of economic reform. It may require higher taxes and, where the social expenditure is of a capital nature and the gains are likely to be shared by future generations, some additional net government borrowing. If sensibly designed, such fiscal devices are superior to restrictions on competition on many counts. They are likely to involve much lower efficiency costs. They can be used to facilitate structural adjustment. They allow the burden of costs and benefits to be more equitably controlled and targeted. And they are generally more transparent and accountable.

There is nothing in the Competition Principles Agreement that precludes governments from spending more money on social assistance (of the passive or active kind). Despite globalisation, governments still have adequate powers of fiscal intervention if they choose to use them. However reliance solely on this fiscal approach to correct the imbalance between efficiency and equity may require a more flexible medium term government stance on revenue, borrowing and spending than is now the case at either state or federal level (Argy, 2001).
Conclusions

The paper has focused on three common criticisms of NCP

- that the democratic legitimacy of the process is questionable;
- that NCP is built on a false premise that increased competition necessarily means increased efficiency; and
- that efficiency is given too much weight relative to other community goals.

Although the subject is complex and does not lend itself to simple generalisations, the paper has expressed certain views which can be summarised as follows.

**Legitimacy of the process.** Concerns about the democratic legitimacy of NCP processes have been allayed as a result of recent changes in *modus operandi*. To go further and give the states much greater discretion would run the risk of destroying the integrity of the cooperative agreement. However there is a need for greater consistency and fairness in the treatment of different groups and sectors.

**Competition and efficiency.** Although it is too early to assess the specific efficiency effects of NCP, economic theory and experience with other competition-promoting reforms both suggest that, provided adequate attention is given to externalities and adjustment costs, the long term winners from NCP will outweigh losers, thus meeting the usual economic efficiency test.

Nevertheless economists who question the efficiency gains of competition policy have two important messages for policy makers. First, they should not start with a universal presumption that competition necessarily improves economic efficiency: this needs to be demonstrated on a reform by reform basis. Secondly, competition policy often needs to be complemented by other policies (for example, to facilitate adjustment and minimise transitional costs) if it is to achieve its maximum potential for efficiency.

**The imbalance between efficiency and broader community goals.** The paper has argued that the present Public Interest Test (PIT) is currently weighted too much in favour of competition and efficiency. It does not give anything like comparable weight to other dimensions of wellbeing such as equity.

Governments may wish to correct this imbalance. If so, one viable response would be to amend the PIT to put all criteria on a more equal footing (although complete equality of treatment would not be practicable). A second possible response could be by way of legislative exemptions. Both these approaches have efficiency costs. A third response would be to smooth undesirable social effects, preferably through ‘active’ social policies. This would be the most effective and efficient way to achieve the desired social outcomes. However the third option is only viable if governments are prepared to give themselves enough medium-term fiscal flexibility to finance the social assistance and adjustment packages.
References


NATSEM (2000), NATSEM News (July).


This paper is an amended version of a talk given to an FMC Conference on the Future of Competition, September 2001. The author is indebted to Gary Banks, John Cosgrove, John Quiggin, Denise Leslie and two anonymous referees for helpful comments, but the views expressed are entirely his own.
Australia-New Zealand Defence Cooperation: Some Considerations

Jim Rolfe And Arthur Grimes

Australia and New Zealand have a long history of defence cooperation (currently under the rubric ‘Closer Defence Relations’ or ‘CDR’) based on treaty arrangements, on shared values and on similar, although not identical, strategic outlooks. (Rolfe, 1995). Each country has always assumed that the other would come to its assistance in times of military threat. At times the relationship has been very close at others no more than correct. No matter the state of the relationship however, until recently there has been a mutual acceptance of the need for each country to maintain a range of combat capabilities for each of the three services within the respective armed forces. The assumption underlying this seems to have been that broadly-based forces at some level of capability would be able to work with each other to each other’s advantage and that between them the two countries would provide the range and quantity of capabilities sufficient to meet any reasonable contingency.

In 2000 each country published reviews giving a framework for decisions on the size and shape of the armed forces for 20 or 30 years (Commonwealth of Australia, 2000; New Zealand Government, 2000). Although Australia explicitly reaffirmed its need to maintain a wide range of combat capabilities, New Zealand’s government argued that ‘a new approach to defence was one of the key policies of the Government when it was elected’ (New Zealand Government, 2000:1). The implication of this was that there could be no guarantees that ‘traditional’ approaches to size, shape and capabilities would continue.

New Zealand did indeed not follow its traditional approach. Decisions have been taken that degrade New Zealand’s combat capacity considerably. Specifically, New Zealand has not taken the option to purchase a third Anzac class frigate (instead deciding to purchase a ‘multi-role’ naval craft), has deferred any enhancement to, or upgrade of, the capabilities of its anti-submarine aircraft, and has removed the air combat force completely from the order of battle. These moves, especially the removal of the air combat force, are a significant change in the previous strategy of ‘wanting a little bit of most things’ and a significant shift away Australia’s position. Army equipment is to be upgraded, but this will do little more than maintain the Army’s capabilities relative to other armies.

Australia’s public view of any New Zealand move to remove combat capabilities (declared before the decisions not to add a third frigate to the fleet and to disband the air combat force) is that it ‘would regret any decision ... not to

Jim Rolfe is Associate Professor, Asia-Pacific Center for Security Studies, Hawaii, USA and Arthur Grimes, at the time this article was written, was the Director, Institute of Policy Studies, Victoria University of Wellington, New Zealand.
maintain at least some capable air and naval combat capabilities’ (Commonwealth of Australia 2000:42). A discussion paper prepared before public consultations on defence requirements noted that to optimise Australia’s forces for ‘operations other than war’ (that is, for humanitarian relief, UN observer missions and the like) would be to take Australia ‘down the road of New Zealand’ (Cheeseman 2000:25). In private, officials and politicians alike are much more scathing of New Zealand’s perceived move to irrelevance in regional strategic matters.

These moves raise questions about the long-standing assumption that each country would act to protect the other. Australia because it may see New Zealand as free-riding and not worthy of assistance unless Australia were directly threatened and New Zealand because it would not have the capability. That outcome would be in neither country’s interests. It may be though, that despite New Zealand’s unilateral approach to force structure issues, both countries can gain. Australia because New Zealand’s decisions can be seen as beneficial to both countries rather than harmful to Australia and New Zealand because Australia would not then see New Zealand as free-riding.

In this paper we consider an analytical framework that could usefully inform policy makers when they consider national force capabilities and the degree to which each country can and should cooperate in determining defence structures. We use an explicitly economics approach to our analysis. This takes us away from normative policy statements of ‘needs’ (which are often ‘wants’) and gives a clear basis for both making and understanding policy decisions. Initially we examine the domestic determinations of an optimal defence force structure for a small country such as New Zealand. The point of this is to determine to what extent national force structure ‘balance’ is feasible and either desirable or necessary. We then consider the international considerations facing any small country and how balance applies internationally. Finally, we consider how the concept of balance could apply between Australia and New Zealand. Underlying all this is an assumption that Australia and New Zealand have sufficient interests in common to make cooperation sensible. If that is not the case, of course, then each country should act unilaterally.

Most of the literature applying economic concepts to defence force structure is written chiefly for and about larger countries (for example, Hartley and Sandler, 1995; Sandler and Hartley, 1995). While conceptually the defence structure issues facing a small country are similar to those facing a large country, the emphasis differs. For instance, a large country such as the United States will inevitably have air, navy and army forces, each with various sub-components (such as strike aircraft, bombers and reconnaissance aircraft within the air force). By contrast, it is inevitable that a micro-state such as Tonga will not have all of these force components (and sub-components).

The emphasis in this paper is on issues facing countries that are not in the micro-state category, but which nevertheless are small. New Zealand, with a population of a little fewer than 4 million people is in this category, and many of the issues it faces are common to other small to medium sized countries, including
Australia. Rolfe (1993 and 1999) gives background on the size and shape of New Zealand’s defence forces and on the policy context within which they operate.

**Balance**

‘Balance’ is important, if only as a concept upon which to hang other assertions. Defence force planners in New Zealand have regularly asserted that balance is essential for the armed forces. The concept as it applies to the New Zealand Defence Force (NZDF) is discussed in New Zealand Government (1991 and 1997). One of the ‘Key Result Areas’ for the NZDF has been ‘the enhancement of balanced military capabilities’ (New Zealand Defence Force, 2000:39). By this the NZDF has meant that New Zealand should maintain as wide a range as possible of conventional military capabilities. For its part, rather than asserting that balance is ‘essential’, the Foreign Affairs, Defence and Trade Committee (1999:56) discussed balance as being ‘not the same as saying that New Zealand needs a conventional balanced force with three separate services’. This latter view has been adopted by the government. Balance, in the government’s view is more to do with having forces relevant to the tasks required of them than it is with having a wider range of forces available for contingencies. In 2000 New Zealand altered its formal position: ‘available forces will be concentrated in areas where they are most needed’ (New Zealand Government, 2000:12). Concentration, the antithesis of balance as the term had been used by previous government, has been manifest by the removal of the air combat force.

Balance, however, is not a well-defined concept. New Zealand’s armed forces were never ‘balanced’ in any objective sense of the word. A small-country’s defence force must inevitably be more specialised (and therefore less ‘balanced’) than that of the United States, if only because it will not have a nuclear capacity. Policy makers must therefore have a clear idea of what balance means and does not mean, in what circumstances it should be aimed for and how it can best be achieved.

Balance is related to the task or tasks required of the forces under discussion. At the highest level there is a need, in virtually all countries, for a defence force of some kind. The ability to defend a country’s own sovereignty in response to particular real or potential threats is generally considered an essential attribute of statehood, and in the worst case may need resort to armed force. A balanced force would be one that could guarantee that security without maintaining more forces than necessary to do so. Regional security (for example, in New Zealand’s case, contributions to the security of neighbouring Pacific Island states and of Australia) may come next in priority, followed by broader ‘foreign policy’ uses of a country’s defence force. This may be either in conjunction with an allied or friendly country (such as with each other or the USA or UK in both Australia and New Zealand’s case) or with a multilateral body (such as the United Nations). Domestic civil emergency uses may also be important. The size and degree of balance of the armed forces should be determined ultimately by the purposes for
Balance may be achieved nationally through the overall structure of the armed forces, or it may be achieved internationally through coordination across countries (for example between Australia and New Zealand). There may, therefore, be a lack of balance at one level and balance at another. For example, it may be sensible for the defence forces to be unbalanced within a country because there is little likelihood that they will operate without cooperation with other countries, but individual services may be balanced so that they can be sent on operations and act more or less independently of support from other countries in certain circumstances. An important question therefore concerns the level at which balance is needed. These questions regarding balance are central to our analysis here.

Small Country Force Structure: National Considerations

Of course, any country has both national and international considerations when it considers an optimal force structure. In New Zealand’s case the conventional wisdom has been that New Zealand has security interests rather than security needs (New Zealand Government, 1991:28). That being so, force structure will not be optimised to defeat any particular enemy or operate in any particular environment. This is different from the case of a country with an enemy of known or assumed capabilities on its borders or a country that, as a matter of policy, has determined that it will, for example, only defend its own shores and have no expeditionary capability at all.

Domestic considerations relating to an optimal structure for the defence forces depend on a number of factors most importantly including:

- the weighting of various objectives set for defence forces; and
- the aggregate funding available for defence forces.

Influence of Objectives on Structure

To show how different broad objectives may influence force structure, assume that there are just two objectives for the defence forces. These may be: ‘pure defence’, relating to sovereignty protection and other warlike operations; and support for foreign policies such as through low-level contributions to the United Nations, or ‘presence’ rather than war-fighting. These two categories reflect the commonly accepted view that defence forces contribute to multiple, jointly-produced outputs (see McGuire, 1995; Murdoch, 1995; Sandler and Hartley, 1995).

If New Zealand had just one force element, the Army say, the two outputs would probably be partially rival and partially non-rival: an increase in resources available for the pure defence roles would increase the resources available to support foreign policy aims, but if resources were to be actually employed in
either role they would not then be available for the other. This introduces the
notion of defence forces being an option. More force gives an option for greater
cortribution to both objectives, even though use for one may crowd out use for the
other at any time.\footnote{1 For New Zealand the concept of option is salient because of
the accepted wisdom, noted earlier, that the country had defence interests rather
than defence needs.}

With two or more forces to consider (the Army and the Air Force perhaps),
the question becomes more complicated. To illustrate this, consider a force with
an Army and an Air Force where both are assumed to contribute equally to the
pure defence function, but only the Army contributes to the objective related to the
support of the country’s foreign policy (the analysis can be extended to more than
two force elements, but this merely adds complexity without increasing our
understanding of the issues). A balanced force (that is, resources to the Army and
the Air Force in the proportion of 50:50) maximises the pure defence objective,
but a completely unbalanced force (that is resources to the Army and the Air Force
in the proportion of 100:0) maximises the foreign affairs objective as the Air Force
does not, by assumption, contribute to that in any case.\footnote{2 It is likely that the
government’s preferences over both objectives will yield a preferred force
structure between the two focused on the Army, but with some Air Force elements
being maintained. The ratio of resources to each could be set at 70:30. Such a
preferred structure is not optimal from a pure defence viewpoint but it is optimal
from a ‘national’ perspective.}

A practical question then arises as to how advice and decisions about defence
force structure and funding should be handled. If, say, the defence forces
contribute to foreign affairs outputs as well as to pure defence outputs, it makes
sense for both foreign affairs and defence officials and other experts to advise on
appropriate force structure to ensure that a range of objectives are met. Differing
policy-making structures, therefore, could have quite different implications for
decisions about force structures. This takes us into the realm of public choice
models (Hartley, 1995) with emphasis on the incentives on each agent (including
lobbyists) to maximise the outputs (or inputs) under their control (Sandler and

In New Zealand this kind of analysis has now been made implicitly (although
probably not in any rigorously analytical manner). This may be seen in the way
the government has expressed force structure preferences based on the ability of
the Army to contribute easily to UN operations whereas the Air Force (particularly
the air combat force) seems to have a more limited role for these operations and
thus has lost its air combat capability.

\footnote{1 This is quite different from the trade-off between resources for defence and say arts
funding: resources used for defence do not give an option for increasing arts outputs
(except possibly for an army band!).}

\footnote{2 By our assumption, any force structure with greater than 50 per cent devoted to the
foreign policy objective results in a deterioration in both pure defence and foreign policy
objectives. Hence the only ‘rational’ allocation of resources is between 50: 50 and 100:0
to the Army and Air Force respectively.}
Influence of Funding on Structure

Consider now a situation where there is just one agreed objective for the defence forces, that of pure defence, (the analysis could be extended to multiple objectives as above, but, again, this would complicate the analysis without shedding light). Assume again two potential force elements within the defence force (the Army and the Air Force) and consider the role of funding constraints on the optimal force structure. Two particular aspects may be important:

- the way force elements are linked to specific defence outputs; and
- minimum feasible levels for either or both forces.

Defence outputs

Consider again the case where defence outputs are a function of the two force elements and examine three alternative situations:

1. We could assume that the total defence output is the sum of expenditures on each of the two forces. In that case both forces should be utilised to the level that additional expenditure on one (within the overall budget) produces a commensurate increase in capability without detracting from the other.

2. We could alternatively assume again that there is expenditure on both the Army and the Air Force but that the defence output is duplicated by the two; neither does anything that the other does not. In this case only one force element (the most productive one in terms of the desired defence objectives) would be employed.

3. A third case would be where the Army and the Air Force supplement each other and one cannot produce any outputs (again, in terms of the desired defence objectives) without the other. In this case, the overall force must be balanced to be viable at all.

These examples show that there is no necessary reason for a ‘balanced’ force but rather that the relationship of the force elements to the defence objectives is important.

With multiple defence outputs (such as pure defence and support for foreign policies) therefore, it is possible that each output will have different requirements. For instance pure defence may require balance whereas contributions to UN peacekeeping may be indifferent to the needs of balance (it may not much matter whether a country contributes an army, navy or air force unit to a multilateral

---

3 Other writers (for example, Warner and Asch, 1995; Sandler and Hartley, 1995) have modelled defence outputs as a function of labour and capital. Since we are primarily interested in force structure, we assume that each force element already embodies an optimised combination (for that force element) of labour and capital and the optimisation calculation is then conducted over the combination of forces.
force provided it contributes something). In a purely rational world, the government would then require Defence to produce a weighted combination of the two outputs with weightings determined by government preferences. In New Zealand’s case though, the history of the usage of the armed forces in the last 50 years and patterns of expenditure on them would seem to indicate that any requirements for ‘balance’ across the armed forces have had less to do with analysis of the kind discussed and more to do with the kinds of ‘bargaining’ described in the public choice models discussed above.

Funding and feasibility

The preceding discussion assumed that there were no minimum levels at which either force had to be maintained in order for it to be operational. In practice this is not normally the case. To take this factor into account, assume that the Army is operable and can contribute to defence outputs at all levels of expenditure, but that the Air Force only contributes once expenditure on it is above a threshold level.

If expenditure on the Air Force is less than the threshold level, the expenditure contributes nothing to defence outputs and is wasted. It would be better to spend nothing on the Air Force and use that expenditure on the Army. But is this optimal? There will be some circumstances where it is and some where it is not. This will depend on whether spending to achieve the pure defence output is higher under the case where there is insufficient expenditure on the Air Force (because the expenditure would be wasted) in which case an unbalanced force structure would be optimal. If on the other hand pure defence is maximised by spending on both the Army and the Air Force, then a balanced structure will be optimal.

The analysis has so far taken aggregate spending as given. However the aggregate level of spending has an important interaction with the effect of the minimum constraint. The more spending there is, the less likelihood there is of the minimum constraint becoming binding.

Consider the policy response to a situation where the optimal size of the Air Force (given a set total defence expenditure level) is less than the required threshold level for effectiveness. There are several policy choices. Firstly, the government could retain the given levels of expenditure and either adopt an otherwise sub-optimal force structure or delete the sub-optimal spending altogether. The choice should depend on the relationship of the Air Force to the Army in the broader production of defence outputs. Secondly, the government could increase defence spending beyond the level it wishes to devote to defence (resulting in a sub-optimal overall allocation of government expenditure) in order to achieve an optimal force structure (viewed purely from a defence force perspective). This indicates a key point: greater defence expenditure not only enables more of the same force elements to be delivered, but also enables more types of force elements to be delivered. The converse also follows.

Sandler and Hartley (1995) argue that declining real defence budgets require a reassessment of force structures both within and across nations. Noting Pugh’s
(1993) analysis that defence costs tend to rise faster than economy-wide costs, they contend (Sandler and Hartley, 1995:116) that ‘there are inevitable downward pressures on the size and formation of armed forces’ and question the wisdom of ‘the long-run trend towards a one-ship navy and a one-aircraft air force’.

This has implications for a small country or, more accurately, for a country with a small overall defence budget. The smaller the defence budget, the more likely it is that the country will adopt an ‘unbalanced’ force. The United States will maintain all available forces and sub-sets of forces. Tonga has virtually none. New Zealand (and Australia), being small but not microscopic, will optimally be somewhere in between — with the exact number of forces being determined by technological and political preference factors. New Zealand, of course, has determined to maintain its defence budget at relatively low levels and delete capabilities rather than to increase the budget significantly and give itself more options.

If, given this analysis, it makes sense in certain circumstances to reduce the number of force elements (as New Zealand has), why does this frequently not happen? One reason suggested by public choice theory is that bureaucratic incentives of decision-makers are such that they will not recommend such a policy (Sandler and Hartley, 1995:117-119; Rogerson, 1995). Sandler and Hartley (1995:117) claim that instead of rationally prioritising across defence elements in relation to equipment procurement, the reality in some countries may be more bureaucratic:

In reality, such decisions are often made on the principle of Buggins’ Turn. Last year, the navy obtained its new aircraft carrier; this year it is the turn of the air force to receive its new fighter aircraft; next year the army can have a new tank!

In New Zealand budgetary considerations are such that expenditure intentions are closely scrutinised. Rarely is there any ability for more than one major equipment procurement at a time. This forces hard decisions and is, in part, the reason for the decisions to reduce the naval combat force and remove the air combat force.4

Of course, if there are three force elements (army, navy, air force) but a single specialised force is optimal (no matter what ideological opinion is held), public choice theory implies that there will always be at least two of the three force leaders who advise against specialisation.

Before finishing this section, we foreshadow briefly discussion of the impact that international coordination may have on the structural issues just examined. The analysis in this section has considered a single country in isolation. There is one aspect in which international coordination could relieve a critical constraint for a small country.

4 Of course, that says nothing about any ideological disposition not to have specific combat forces.
Recall the case where the country requires both (or all) forces if any defence output is to be produced. If the country can rely on an ally (either formal or informal) for provision of one force, then the appropriate production function should include the other country’s input of that force to the domestic defence effort. This could result in the adoption of an unbalanced rather than a balanced force structure in the smaller country even where an analysis of defence outputs produced by each service would mandate a balanced force in the absence of international cooperation.

**Keeping the Options Open**

So far our analysis has been entirely static. However in an uncertain environment, once the time to bring forces to an operational capability is taken into account, the analysis becomes dynamic and further important complexities are introduced. In these circumstances it may become more costly to specialise because the loss of capability arising from the need to bring mothballed forces to readiness must be taken into account.

The impact of this factor on structural choices is, however, complex. While reducing the benefits of specialisation compared with the static case, the need for some specialisation by a small country remains. Quite evidently, the need to retain options is not an argument for a small country to maintain all defence elements maintained by the United States (submarines, nuclear missiles and the like).

The force elements maintained by a small country may in the real world, however, be greater in number than would be the case under a static (and probably) theoretical framework and/or will tend to be more ‘generalist’ in nature. As an example, if an Air Force is to be maintained, its structure may differ once dynamic factors are taken into account. A static framework with a given environment and given objectives may result in choice of one type of aircraft to carry out a specific function (such as interception of enemy aircraft). However once environmental or other uncertainties related to the range of possible tasks are taken into account, the choice of aircraft may be modified to a model which is capable of carrying out a broader range of functions (both ground attack and interception for example), albeit with a lesser ability to carry out its currently perceived main function, or indeed any of the functions.

A corollary of this analysis is that force elements capable of being used across a broad range of environments and to meet a broad range of objectives will be favoured over force elements which can only be used for narrower tasks. This factor could conceivably result in retention of a specialised defence force (in terms of fewer force elements) if the force elements retained were capable of being used across a broader range of defence outputs (or environments) than could a more ‘balanced’ force structure. Again, balance is related to the tasks that can be carried out rather than to any specific structure adopted.

It is possible though, that a balanced force would so run down the resources available to the force element appropriate to a specific environment that the
defence output in that environment would be less than the defence output under a specialised force structure and less that sufficient to be effective. In this case, even though there is environmental uncertainty and we know that in one environment a balanced force will be preferable to a specialised force, the optimal force structure is still a specialised force.

The key result, then, is that environmental (or objective) uncertainty does not automatically lead to the favouring of a balanced force even where a balanced force is preferable under some environments. One must still examine the opportunity costs under alternative environments (or objectives) of maintaining a greater number of force elements.

**Small Country Force Structure: International Considerations**

We have taken as given that a government will have a number of (possibly prioritised) objectives for its defence forces. We also take it as given that a *sine qua non* of any defence force is to protect the country at least from low-level threats to the country’s sovereign integrity (such as terrorism or unauthorised incursion into a maritime country’s EEZ). Beyond this, a country’s optimal defence structure will depend on the national factors discussed above and on a range of international considerations such as the extent and nature of its alliances, and on regional and broader foreign policy objectives to which defence forces might contribute.

This range of potential objectives suggests that decisions about optimal defence structure cannot, in general, be made purely by reference to country-specific factors. This is so unless the only objective set for the defence force is protection of a country’s own sovereignty and even then only if such protection does not involve cooperation with other countries. Once the objective set is extended beyond this, a country’s armed forces will necessarily interact with forces of other countries. Hence questions of structures to optimise performance in an international environment must be considered.

For New Zealand, the relationship with Australia is fundamental: ‘there is no strategic relationship closer than that with Australia’ (New Zealand Government, 2000:6). The relationship is underpinned by both treaty and explicit policy statements, the gist of which are that an attack on either country would be considered to be an attack on the other. The treaty and interest based relationship between Australia and New Zealand does not, however, require either country to conform to the other in terms of either strategic posture or force structure and capabilities. It does suggest though that close coordination should occur (as it does) and that, at the very least, both the capabilities and interests of one country should be taken into account by the other when decisions are being taken. But the optimum outcome is not necessarily clear.

Economic theories of military alliances extend back to the work of Olson and Zeckhauser (1966) who discussed ‘deterrence’ as a public good and who noted that all lose when defence spending decisions are determined non-cooperatively. But small countries are likely to free-ride in such a situation by ‘under-spending’
on defence forces which contribute to mutual deterrence. Australia has ‘perceptions — inaccurate or otherwise — of an imbalance of responsibilities that each country bears within the bilateral relationship’ (Rolfe, 1999:71). Indeed analyst Paul Dibb (2000) has described New Zealand as being a ‘strategic liability’ not a ‘cooperative ally’. Even if political coordination across allies prevents such free-riding, choices for a small country differ from those of a large country in terms of whether a country can take other countries’ choices as given. In many policy spheres, a small country can take foreign decisions as exogenous and then optimise its own decisions given this environment.

In regional defence where the threat to all members is likely to be similar, however, the approach may (or should) be different. Defence partners (such as New Zealand and Australia) routinely discuss appropriate force structures between themselves, even if they come to no agreement on what is necessary and what is sufficient. A number of justifications may be used for such an approach. Firstly, there is the issue of interoperability (consistency of equipment and processes across forces). As Sandler and Hartley (1995:223, 339) note, if the forces are to work together then a high degree of interoperability is likely to be beneficial. A caveat to this is, of course, where one country (likely to be the smaller) is locked into purchasing inferior or more expensive equipment mandated by the regional partner, perhaps on the basis that the equipment is produced in that country. However, this consideration could be handled via the traditional ‘small country’ approach: the small country observes or is informed of a larger partner’s choices and then takes these as given in making its own decisions. Interoperability per se is not a strong argument for cooperation in determining or influencing structures across the two countries.

A second and better reason could be that a given total defence effort of the two countries could be achieved at lower cost through coordinating their respective structures than if the two countries were to make independent decisions. This reflects an economies of scale argument; defence costs and/or outputs with international coordination may differ from those without coordination.

Take a two country case. One country (Australia, say) is large, and the other (New Zealand) is small. There are two decision-making possibilities: independent and coordinated. In the independent case, New Zealand can observe and take Australia’s decisions as given and choose its own defence structure accordingly. Australia must make its choices without necessarily knowing New Zealand’s choice. Without coordination, an inefficiency will result. If Australia for some reason cannot count on New Zealand making a sensible decision, or cannot count on it maintaining its ‘fair share’ of expenditure, Australia would have to maximise its own defence output to achieve its own needs. This would result in a ‘balanced’ force viewed purely from Australia’s perspective, but not necessarily a force balanced for regional security needs. The same is true for New Zealand. If New Zealand takes the resulting Australian force as given, it will find that elements of its own forces will be different from their optimal coordinated level. Depending on the size of the fixed cost and the nature of the production function, it may be optimal (if making decisions independently) for New Zealand also to
maintain a full range of force elements, at the expense of additional cost and loss of regional balance. This seems to be the position today.

In the coordination case, Australia and New Zealand would cooperate to achieve a common defence objective. If the two countries agree on how much they are individually willing to spend on defence and can coordinate policy, they will maximise their combined defence outputs through fixing the size and capabilities of their respective individual services at appropriate aggregate levels, and ensuring that the structure of forces across the two countries minimises unnecessary fixed costs. With one large and one small country, this is likely to result in the larger country having all capabilities to some level and the smaller country specialising. This is the familiar ‘comparative advantage’ result discussed in much of the international trade literature. Sandler and Hartley (1995:39) also note that comparative advantage analysis may lead each country within an alliance to specialise in providing sub-sets of security outputs. This results in a balanced force across the two countries. Both countries gain from the adoption of such a force structure compared with the independent decision-making case.

This gain exists even though it results in an ‘unbalanced’ force (from a single country perspective) in New Zealand which no longer has a particular capability. It also results in an ‘unbalanced’ force for Australia even though that country has a full range of capabilities. To see this, note that elements of all force elements will be provided by Australia, while New Zealand will provide a larger proportion of one force element (or a few) than its size dictates; thus Australia has a larger ratio of those forces not provided by New Zealand than it would if it were taking a force structure decision in isolation.

Clearly, independent decision-making will be sub-optimal relative to the coordination result (if total expenditures are identical). To see this, note that either country may independently decide on the following options:

1. To adopt a force structure with a full range of force elements.
2. To adopt a limited and specialised force structure.

Option 1 must be inferior to the coordination result since an extra fixed cost is incurred which must lower total defence output for any given level of expenditure compared with the coordination outcome.

Option 2 must be inferior to the coordination result (unless that result is achieved by accident), even though the fixed cost is not incurred, since some elements across the two forces will be below (and others above) the optimal level. Hence the combined defence output is not being maximised.

Importantly therefore, because of the existence of fixed costs, a small country can maximise its contribution to a multi-country defence effort by adopting a small number of well-prepared force elements; the larger the country, the more force elements it will adopt. Essentially, this is a formal derivation of the conjecture by Sandler and Hartley (1995:224) that if NATO members were certain that in the event of an attack their allies would respond, each nation would have an incentive to specialise in its force structures rather than create a totally
independent capability (each nation with an army, navy, air force). A caveat to
this result is that it relies on each country having identical objectives and having
certainty about the responses of its allies in circumstances where combined
responses are sought. Without identical objectives or certainty, the case for
specialisation diminishes — but this is because of a lack of common objectives or
certainty; not because of a supposed optimality of each country contributing
‘balanced’ forces to a common defence force. The case for specialisation also
diminishes if a government, for political reasons, is committed to maintaining an
independent defence capability (Sandler and Hartley, 1995:226).

At the moment Australia and New Zealand make their decisions more-or-less
independently of each other. The result is, therefore, sub-optimal in terms of the
countries combined needs, assuming they have both certainty in the other country
and more or less identical objectives.

A Specific Case: The New Zealand Air Combat Force

The Army has been the main defence force element contributed by New Zealand
to international operations since World War Two. At times New Zealand has also
contributed ships (for example, frigates to the Multinational Interception Force in
the Arabian/Persian Gulf) and transport aircraft (especially to carry humanitarian
relief). Strike aircraft have been a notable absentee from New Zealand’s
contribution to international forces over the past three decades despite New
Zealand having two A4K Skyhawk squadrons. Strike aircraft were part of New
Zealand’s commitment to the Commonwealth Far East Strategic Reserve in the
1950s and 1960s, and were used operationally from 1955 to 1960 during the
Malayan Emergency and deployed in 1964 during Confrontation with Indonesia.

Despite the lack of operational usage, the air combat force was maintained
because of its option value. In 1997 the government (New Zealand Government
1997), set out an investment path for the NZDF consistent with an explicitly stated
policy of maintaining a ‘balanced’ defence force. Balance was most obviously
demonstrated by defence expenditure which in 1999 (after abstracting from
headquarters expenses) was allocated 37 per cent to the air-force, 33 per cent to
the army and 30 per cent to the navy (NZDF, 1999b:19), but was also seen in the
maintenance of core combat and support elements for each of the services.

Included as a key element of the investment policy was the intention to ‘lease 28
F16A/B aircraft. The aircraft are expected to arrive in early 2001 and will replace
the 19 A4K Skyhawk aircraft’ (NZDF, 1999b:18).

A change of government in late-1999 saw a reassessment of this policy. The
new government was less wedded to the idea of New Zealand maintaining a
balanced defence force, wishing to focus on optimising New Zealand’s defence
contribution, especially in international peacekeeping roles. Following
establishment of a committee to investigate the F16 lease deal (headed by a former
Defence Minister Hon Derek Quigley) which reported favourably on the deal, the
government decided, in March 2000, not to proceed with the F16 lease. Prime
Minister Helen Clark (2000) stated that while ‘the arrangements offered by the
United States for the F16s were good ones’, the F16 acquisition was worthwhile only ‘provided that expenditure now on upgrading the air combat force was a priority and that the Defence Force’s other priorities were also affordable’. She concluded that this was not the case. In particular (Clark, 2000):

there is no priority setting mechanism in the New Zealand Defence Force which effectively and consistently links individual activities or projects to the government’s most pressing national security concerns.

This observation indicates that the new government did not consider the previous matching of resources to objectives to be adequate (that is, the considerations listed earlier in this paper had not been adequately addressed). In background analysis of the F16 deal, Grimes (2000) demonstrated some of the trade-offs incurred in New Zealand maintaining an air strike capacity. That analysis assumed that New Zealand could not realistically scale down the existing capacity; that is, it assumed that an air combat force of 19 aircraft is the minimum required level for the force to exist in a meaningful fashion. Thus the policy choice (consistent with the discussion above) is either to discontinue the force or to keep it at least its current capacity. If the force were discontinued and if aggregate defence funding were maintained constant, the analysis indicates that:

- ready reaction and other army forces could be increased by at least 30 per cent; or
- at least an additional frigate could be added to the (3-frigate) navy; or
- maritime patrol forces could be more than doubled.

The air combat force has been disbanded and the government has taken decisions to increase total defence expenditure slightly and to substantially bolster the army’s equipment but not to materially increase that of either the navy or air force, although patrol craft and a multi-role vessel for the Navy will be acquired.

While the official Australian reaction to New Zealand’s decision to remove the air combat force has been largely negative, in part because a key role for New Zealand’s strike aircraft included its ‘contribution to the defence of Northern Australia’ (Wilson-Roberts, 2000; NZDF, 1999b:14) and in part because of a more generalised belief, alluded to earlier, that any diminution of New Zealand’s defence capacity is in itself a threat to Australia.

Grimes (2000) also demonstrates that ready reaction and other army force expenditure is much more labour-intensive than is air combat expenditure. The ratio of labour:total expenditure for the ready reaction and related forces is 49 per cent compared with 28 per cent for air combat forces; capital:total expenditure ratios are almost reversed. Given a public policy desire to increase training and employment of lower skilled workers, and a desire to cut imports, a transfer of expenditure from combat aircraft to army forces therefore potentially results in other public policy benefits.
From our analysis above, the Australian objection possibly rests on faulty analysis. Australia, seven times New Zealand’s size, has a sizeable air combat capability. If the two countries could coordinate their defence efforts, it is likely to be optimal for Australia to enlarge its air strike capability by the 20 or so combat aircraft currently retained by New Zealand (reducing its other forces commensurately), leaving New Zealand to enlarge its ready reaction and other army forces by over 30 per cent so as to contribute more strongly to situations such as in East Timor and Bougainville. It appears that, in practice, the New Zealand decision is leading to an increase in the capacity (if not the total numbers) of Australia’s air combat force as most of New Zealand’s air and ground crews are moving to join the Royal Australian Air Force, thus relieving shortages in that force.

Conclusions

A small country must decide not only the scale of its armed forces, but also their scope. Inevitably, the smaller a country’s defence expenditure, the smaller will be the number of force elements which it can maintain at credible operational levels. Factors which will influence the choice of force structure for a small country include: the nature of defence objectives; the nature of international cooperation amongst national defence forces; the overall level of funding for the armed forces in conjunction with technological factors including the size of fixed costs for each force element; and the degree and nature of environmental uncertainty.

As we move down the continuum of country size, a country must, if it is to establish an optimal force structure, choose to forego defence force elements that would be considered an integral part of the defence force of a larger country. This does not necessarily signify a lack of commitment to defence on the part of such a small country. If anything, it signifies the opposite: a country that is willing to increase its defence contribution by reducing its number of force elements is showing an increased commitment to defence outputs (but not to certain defence careers) given limited resources.

Discontinuing existing force elements, even where this is the optimal path, is subject to potentially severe public choice objections. ‘Informed sources’ — both domestic and international — may oppose an optimal reprioritisation of defence expenditure, especially where this involves the scrapping of one or more force elements. Governments must then make a decision as to whether to follow a path of optimal resource allocation or a path of least resistance.

Experience suggests that the latter path often predominates. From New Zealand (and other countries’) experience, it is likely to take a brave government to decide explicitly to reduce the number of force elements. A more likely scenario (if the ‘buggins turn’ approach is not adopted) is a long-term run-down in capability of one or more force elements. While this approach retains an option to revitalise that capability in future (more cheaply than if it had been totally discontinued), it is only a second best alternative to discontinuation where the latter is optimal. The latter approach may enable substantial savings in fixed costs.
that could be allocated to enlarge the contributions of other existing force elements. Therefore a choice of defence structure which may on the surface appear the least supportive for defence (that is, discontinuation of a force element) could prove instead to be a significant contributor to bolstering defence outputs.

This analysis has shown that although New Zealand has apparently reduced the capacity of its armed forces, in practice it might well have increased not only its own overall utility (because proper attention will be paid to the Army which will become more effective) but also that of elements of Australia’s. If that is the case, then there is considerable scope for the two countries to consider other areas for the armed forces to complement each other (in counter-terrorist capabilities perhaps, or for the defence of northern Australia) and for the total defence effort of the two countries to be enhanced. This would mean a degree of specialisation, more so by New Zealand than Australia. But given that each country is more or less specialised in any case, this would make explicit what is already implicit. A greater barrier is likely to be the question of ‘sovereignty’ and independence. If the two countries join their armed forces either functionally or at a higher political/constitutional level there will be cries that one or other country has relinquished sovereignty and independence of action. The question of trust and free-riding will also be raised, especially by Australia. These are serious issues, but they are capable of being resolved.

There are advantages to be gained if the two countries could cooperate in their defence planning. There are some caveats of course. Each country would have to be certain of the other’s good faith and they would have to be able to agree on a common set (and understanding) of defence objectives. These should not be insurmountable problems. The two countries are closer in outlook and capacity than almost any two others. The two countries already have an economic relationship that is getting ever closer and where that economic relationship has already gone is suitable ground to consider the destination for the defence relationship.

References


Rolfe, J. (1999), The Armed Forces of New Zealand, Allen and Unwin, St Leonards, NSW.


This work grew out of research conducted for and funded by the New Zealand Treasury. We are grateful to the Treasury for their support. The views expressed herein should be attributed solely to the authors and not to their respective institutions or to the New Zealand Treasury. We thank Ross McLeod and four unnamed referees for their very helpful comments on earlier drafts of this paper. The interpretations and any errors of fact or omission are of course the responsibility of the authors alone.
Brain Drain in New Zealand: Issues, Evidence and Implications

Wai Kin Choy and Hayden Glass

There is public concern that New Zealand is experiencing a ‘brain drain’ — the loss of a large number of its brightest citizens to other countries. Some of the related public policy issues are discussed in this paper. The paper begins with an outline of possible reasons for concern about a brain drain. This is followed by a summary of the available evidence on whether New Zealand is actually experiencing a brain drain, examining the numbers of emigrants and immigrants, and their skill and age compositions over time. The focus is on the aggregate picture, rather than specific occupational categories. More detailed discussions of the evidence can be found in Glass and Choy (2001). The paper also presents some thoughts on the related policy implications.

When a Brain Drain Is of Concern

A brain drain refers to the numbers of emigrants and immigrants (that is, a brain drain is a net outflow of people), or to their attributes (that is, a brain drain is a net outflow of the high-skilled or the highly-experienced). The latter tends to be the interpretation normally adopted, and is the focus of this paper.

A net outflow per se may be of public policy concern if one believes that a certain critical mass is needed to support the nation’s development and progress, that is, if a larger population is needed to achieve economies of scale and higher rates of economic growth and improved living standards. The problem lies in determining what the optimal size is.

Net outflows may not be the only cause for concern. If the people leaving and entering the country are different in some relevant way, there may still be concerns even if there were no net outflows. Similarly, a country may not be better off just because there are net inflows (say, there were zero population growth, all the people leaving were high-skilled and all the immigrants were low-skilled). Therefore, the composition of both the inflows and outflows matters. A net loss of skilled workers is equivalent to a reduction in the domestic human capital stock and may be of concern.

The key link between living standards and the skill composition of migration flows is through income — a central component of living standards. The key to
higher per capita incomes is productivity. One crucial contributing factor to higher productivity is skilled workers (see Claridge and Box, 2000). They are better at creating, improving and applying new knowledge, which stimulates technological advance and thus productivity.

The impact of changes in the skill composition of the labour force on production and productivity depends on the degree of complementarity between skilled labour and the other inputs of production — capital and unskilled labour. The impact of a brain drain on productivity depends partially on whether skilled workers compete with unskilled workers (are ‘substitutes’ for them), or enhance their productivity (are ‘complements’ for them).

There is empirical support for complementarity between skilled labour and capital (Krusell et al., 1997; Golden and Katz, 1998). However, evidence on the degree of complementarity between skilled and unskilled labour is mixed (see Hamermesh, 1993). To the extent that they are complements, a reduction in the number of skilled workers, as would occur with a brain drain, would lower the average productivity of the economy and hence the earnings of less skilled workers and the return on capital. It would also have the effect of raising the productivity of remaining high-skilled workers. The overall effect on productivity depends on the degree of complementarity and the shares of high- and low-skilled workers in the labour force. It is possible that a brain drain could be harmful for the source country in aggregate (see Bhagwati and Hamada, 1974).

The ‘new growth’ literature suggests that the negative implications of a brain drain on the source country are magnified. For example, Hague and Kim (1995) posit that the emigration of people with high levels of human capital reduces the growth rate of the effective human capital that remains in the economy, and thus generates a reduction of per capita growth in the source country.

Under this scenario, not only is high-skilled labour complementary with capital and low-skilled labour, but high-skilled workers enhance technological innovations and their diffusion. There could also be significant learning-by-doing. The main implication is that a continuous outflow of high-skilled labour would deplete the source country’s level of human capital and thus reduce that country’s capacity to achieve as much technological progress as other economies.

Both the standard and new growth models have recently been challenged in various theoretical papers that examine the impact of migration on human capital formation within a context of rationed migration flows (see Beine, Docquier and Rapoport, 2001; Mountford, 1997; Stark, Helmenstein and Prskawetz, 1997). These theoretical papers assume that workers weigh the costs of acquiring skills against prospective market rewards for enhancing skills, both at home and abroad, and make optimizing decisions. The possibility of migration changes the opportunity set and the incentive structure. In particular, the assumed higher returns to education abroad create an incentive for the population to upskill and increase their human capital. In the extreme, if there are large numbers of people trying to compete in the international labour market, then there will be an increase in the level of education acquisition in the home country. Given that only a fraction of the educated residents will emigrate (via some rationing device either
from the destination country, or by the source country), the average level of education of the remaining population in the source country would increase. This is known as the *ex ante* ‘brain gain effect’.

The argument above depends on there being some rationing. Where there are free flows of people, the scope of this mechanism can be greatly reduced or even eliminated. For example, due to the lack of rationing under the Trans-Tasman Travel Arrangement (TTTA) between New Zealand and Australia, it is unlikely that workers in New Zealand who upskill themselves in the expectation of higher returns to their skill in Australia will actually remain in New Zealand. However, for other country destinations — where rationing still applies — the *ex ante* brain gain effect may hold.

The crucial question then is which effect dominates — the *ex ante* brain gain effect (migration prospects fostering investments in education because of higher returns abroad) or the *ex post* brain drain effect (that is, educated agents actually migrating). If the first effect dominates, then the average level of human capital is higher in the economy open to migration than in the closed economy. In such circumstances, there will be a ‘beneficial brain drain’ (see Beine, Docquier and Rapoport, 2001; Mountford, 1997). This is an empirical question to be determined. Unfortunately, there is relatively little empirical work on this (see Carrington and Detragiache, 1999:7).

If there is significant return migration, then the loss of highly trained personnel at any one time can have beneficial impacts on the source countries in the medium to longer run. This is why some people have labelled this kind of scenario a ‘brain circulation’ rather than a ‘brain drain’ (for example, Johnson and Regets, 1998). The level of productivity and hence, the national welfare may rise if those skilled workers who went abroad return to the home country with improved productivities. The critical question then is whether there is in fact return migration and what the profile of migrants returning to the source country looks like. Even in a situation of ongoing net outflows, if the returnees are more skilled than those leaving, the net economic impact on the country might not be negative.

In short, a drain of brains and a welfare gain need not be mutually exclusive. In fact, a drain of brains can be the very cause of a welfare gain if it stimulates human capital acquisition in those that remain, or leads to higher-skilled return migration.

But it is not all good news. Based on the theoretical literature, there would be cause for public policy concern if there were large net outflows of the high-skilled

---

1 A referee argued that it is difficult to see how a rationing of migration opportunities abroad would necessarily raise education levels in the home country, *unless the rationing was unforeseen*. If the rationing were anticipated, it would probably reduce the expected returns to tertiary education, and thus reduce the *ex ante* brain gain effect. But because of the pervasive uncertainty about the returns to skills and education, either an under-investment or over-investment (from the individual’s perspective) could still occur. Having said that, this point is merely to illustrate that it is *possible* for large migration flows at a particular point in time to benefit the source country.
over the long term, and there was no rationing of emigrant numbers to other countries, and little return migration.

Those in the working age population are usually assumed to be net contributors to the economy (particularly in the labour market). Such contributions help determine the standard of living for the average person in the population. Therefore, even if there are net outflows, one needs to look at the age composition of the flows. Departures of those who do not fall into the working age category arguably would not raise as much economic alarm as if there were large departures of working age people.

Age composition is also important as a proxy for the amount and breadth of work or business experience that a migrant has. For this reason, from a public policy standpoint, we should be more concerned if we are losing the older group from the working age population than the younger groups.

There are numerous social impacts from emigration and immigration (see New Zealand Institute of Economic Research, 1994; and 1996). For example, changes in ethnic and cultural diversity can influence how current New Zealanders adjust to new residents and citizens and the changes they bring to national identity. There could be a lower critical mass for maintaining existing New Zealand culture, resulting in change in national identity and culture.

A brain drain could also have significant negative fiscal effects for the remaining population including through:

- **Education investments** — Most education services are substantially funded by taxpayers, who therefore have an interest in getting a return on that investment in the form of taxes from the educated person over his/her life. To the extent that educated people take their accumulated education overseas with them, the implicit contract with taxpayers is thwarted. New Zealand taxpayers end up, in effect, subsidising the growth of other countries. The effect is stronger to the extent that the educated can avoid paying their contribution by avoiding taxes or student loan repayments.

- **Other fiscal costs** — New Zealanders qualify for subsidised health care by birth or by securing permanent residence. They can qualify for superannuation by working or living here for a minimum period. There is a risk that New Zealanders will go overseas and avoid the tax that could be expected to fund these costs, and return to New Zealand for health care or for superannuation at the cost of those who remain behind.

These non-economic impacts could be significant. But, as with the concerns about the numbers of flows per se, their implications for public policy on the brain drain are inconclusive.

This section has covered various reasons why public policy might be concerned with a brain drain. Broadly speaking, policy makers should be concerned if:
• there are net outflows of the high-skilled and no rationing exists in the destination country (because then the \textit{ex ante} brain gain effect does not apply);
• there are net outflows of the highly-experienced (those in the middle or late working age population groups); and
• there is limited return migration.

The Evidence

This section summarises very briefly the evidence on the numbers, skill composition and age of migrant flows to and from New Zealand (see Glass and Choy, 2001 for more detailed evidence). We consider migration flows between New Zealand and the rest of the world (all countries, including Australia), and the flows between New Zealand and Australia.

The primary source of international migration data is the arrival and departure cards. The arrival card records, \textit{inter alia}, how long people have been away from New Zealand (for residents) or how long they intend to stay in New Zealand (if they are not residents). The departure card records how long people intend to be away from New Zealand (for residents), or how long they have stayed (for non-residents). From their answers on these cards, people are classified as either Permanent and Long-term (PLT) Migrants, who were away or intend to stay away from their home country at least a year; or Short-term Visitors, who were away or intend to stay away less than a year. The main caveats on the use of this data have already been covered extensively in the literature (see Glass and Choy, 2001; Carmichael, 1993a; and 1993b).

Figure 1 shows permanent and long-term (PLT) migration flows to and from New Zealand between 1961 and 2001. The general New Zealand convention to analyse the data using years ending March is followed throughout this paper (see Bedford and Bedford, 2001 for an update on the 2001 September quarter international migration statistics). The net total migration flow is shown alongside the net PLT series. The net total series is one way of accounting for changes in people’s intended duration of stay (see Bedford, 2001 and Glass and Choy, 2001).

While there are net PLT outflows recently to the rest of the world, they are small, cyclical and follow a long period of inflows. In the longer term (for example, 10 years or 40 years), New Zealand \textit{gains} people from migration. The short-term statistic (net PLT) can give a misleading picture. Accounting for changes in people’s intentions (net total) seems to suggest net inflows recently. Given this volatility in short term figures, it is important to view migration over the medium to long-term.
Although not shown in the figure, over the last 40 years, there are net outflows of New Zealand citizens almost every year. However, these departing New Zealand citizens have been replaced by slightly larger inflows of citizens of other countries. Therefore, to date New Zealand has experienced a replacement migration or exchange, rather than a ‘brain drain’.

On a per capita basis, contemporary outflows (and inflows) of New Zealanders are high by comparison with countries of similar size (for example, Ireland, Singapore and Norway) — there is more ‘turnover’ in the New Zealand population than in these three countries. For example, over the period 1996-1998, in the case of Ireland, gross migration inflows were approximately 1.1 per cent of the total population, and gross outflows about 0.7 per cent of the population, giving a net migration rate of 0.4 per cent. In the case of Norway, the gross inflows and outflows rates over the same period were 0.8 per cent and 0.55 per cent respectively, resulting in a net migration rate of 0.25 per cent (see OECD, 2000). The corresponding figures for New Zealand are 2.0 per cent, 1.5 per cent and 0.5 per cent respectively. One can argue that the high levels of immigration under the points system in New Zealand (high on a per capita basis, again by comparison with Australia, and the US, and similar to Canada) coupled with the significant PLT outflows of New Zealanders, and the persistent return flow of around 20,000 to 22,000 per year, is a clear indication of an exchange rather than either a drain or a gain.

As discussed in more detail in Glass and Choy (2001), there is only relatively unreliable information on the skills of migrants. This is based on the question on the departure and arrival cards that asks for travellers’ occupations. In over half
the cases, there are no data on the occupations of the migrants (there is no requirement to fill in the card completely), and consistent data are available only since 1992. In addition, occupational data are available only for those who identify themselves as PLT migrants. One should bear these limitations in mind when making inferences using such data.

This analysis places occupations into one of three skills levels — high-skilled (New Zealand Standard Classification of Occupations (NZSCO) codes 1 to 3), semi-skilled (NZSCO codes 4 and 5) and low-skilled (NZSCO codes 6 to 9) (see Shevland, 1999). High-skilled includes professionals, associate professionals, technicians, and managers. The semi-skilled group includes clerks, service and sales workers, and the low-skilled category includes people involved with routine production work, and manual labour. The meaning of these terms is not exact, and the terms are used for simplicity, rather than for precision. Some relatively skilled tradespeople (NZSCO code 7) are included in the low-skilled grouping (Humphris, 2001).

Figure 2 compares the three-year average percentages of PLT departures and arrivals in each skill category at the beginning and end of the 1992-2001 period. It is noted that these are percentages of all those who specify an occupation that could be coded.

**Figure 2: Skill Distribution of PLT Migration Flows 1992-2001**

![Figure 2: Skill Distribution of PLT Migration Flows 1992-2001](chart.png)

(3-year averages at start and end of the 1992-2001 period)
Figure 3 below shows how the skills of people migrating to the rest of the world compare with those of the remaining New Zealand population for the 1997-1999 period.

The data available on occupation seem to suggest that the skill pattern of departures to the rest of the world has not changed much over the last decade, whereas arrivals are getting more skilled, at least in recent years (see Figure 2). Departures to the rest of the world appear to be more skilled than the remaining population (see Figure 3). However, these skill-biased departures to the rest of the world have to be considered alongside the high proportion of high-skilled people immigrating to New Zealand. In fact, there is some evidence that recent arrivals have contained a greater proportion of people in the ‘high-skill’ category than have departures. In short, there seems to be a brain exchange, rather than a brain drain.

**Figure 3: New Zealand’s Skill Distribution 1997-1999**

![Skill Distribution Chart]

In addition, there is a substantial level of return migration, which could reduce concerns about a brain drain. Since the early 1980s, more than 20,000 New Zealand citizens have been coming back to the country each year (see Lidgard, 1993, 1994 and 2001).

Data on age of PLT migrants seem to suggest that the age of both those departing and arriving have increased. More disaggregated data reveal a drain of younger New Zealanders and a gain of adult non-New Zealand citizens, hence the label ‘exchange young people for adults’. Controlling for changes in the age distribution of the remaining population, age-specific emigration rates further support the point that departing New Zealanders have grown older. In particular, the likelihood of emigrating for 15-19 year olds has dropped, while the chances of leaving for 20-29 year olds have increased.
Trans-Tasman flows are different from those to the rest of the world. A large number of people (predominantly New Zealand citizens) are going westward, and these net outflows are volatile (see Figure 4).

**Figure 4: PLT Migration Flows to and from Australia 1947-2001**

The flow of New Zealand citizens to Australia is not biased toward the high-skilled. Instead, it is fairly representative of the remaining population in New Zealand (see Figure 5), hence the label *same drain*, rather than a *brain drain*. This evidence is consistent with the existence of a common labour market between the two countries. The main effect of the common labour market has been quite different to a brain drain. Rather than draining only the high-skilled, it has allowed the migration of a broad mix of people, including low-skilled and semi-skilled New Zealanders.

**Figure 5: Skill Distribution of New Zealand Citizens 1997-1999**
Figures 3 (NZ-Rest of the world) and 5 (NZ-Australia) are not directly comparable because Figure 5 focuses specifically on New Zealand citizens, while Figure 3 looks at all departures, regardless of citizenship. However, our conclusion that departures to the rest of the world are skill-biased but those going to Australia are representative of the remaining population is not affected (see Glass and Choy, 2001).

Birrell et al. (2001) present a different assessment of the evidence in relation to the skill composition of trans-Tasman flows, noting that New Zealand is the largest source country of gain in professionals to Australia. This paper makes somewhat different points from Birrell et al. It looks at the skill mix of the entire flow, while Birrell et al. look at the (large) number of professionals in the flow from New Zealand to Australia (not the share of the flow that are professionals). In addition, the way our data set has been put together differs from that of Birrell et al. This paper focuses on New Zealand citizens, whereas Birrell et al. use New Zealand as a country of last or next permanent residence, and hence they include, for example, Australian citizens and third country citizens who have resided, or are intending to reside, in New Zealand. Another difference is that this paper focuses on PLT movements only, but Birrell et al. analyse total movements (that is, including short-term flows).

Policy Implications

Table 1 summarises the key pieces of evidence about the numbers, skill and age compositions of migration flows.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>NZ-Rest of the World*</th>
<th>NZ-Australia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net Numbers</td>
<td>Inflow</td>
<td>Outflow</td>
</tr>
<tr>
<td>Skill</td>
<td>Brain exchange</td>
<td>Same drain</td>
</tr>
<tr>
<td>Age</td>
<td>Exchange young people for adults</td>
<td></td>
</tr>
</tbody>
</table>

*Rest of the world means all countries, including Australia.

A key issue is not whether migrants are more highly skilled, but whether they integrate well enough into the labour market to be able to make use of those skills. The unemployment rates of recent migrants are typically high. Winklemann (2000) gives an overall rate of 35% for migrants in the first year of residence in New Zealand based on 1996 Census data. Rates were substantially lower for younger age groups, and those from English speaking countries, and up to 59% for
migrants from South Asia. Similar results are cited in Bedford, Ho and Lidgard (2000) in relation to specific ethnic groups.

Unemployment rates for new labour market entrants are frequently very high. The key question is what happens over time. Poot, Nana and Philpott (1988) present evidence that the likelihood of immigrants being unemployed decreases as time in New Zealand increases. The income of the overseas born who had been in New Zealand 10-14 years in 1981 could be ‘favourably compared’ with the income of the New Zealand born. Pacific Islanders were particularly disadvantaged on arrival, but tentative evidence suggested that they experienced rapid declines in unemployment and increases in income over time. The explanation for the differences between overseas and New Zealand born focused on skills, and particularly on English language ability.

Results from Winkelmann and Winkelmann (1998) take this point significantly further. Their findings indicate that some immigrants have difficulty integrating into the labour market over time — particularly those from Asia or the Pacific who do not speak English. A typical immigrant, despite being relatively highly educated, was likely to have a lower income and lower probability of participation and employment than a New Zealand-born person of the same age and education level in the first few years after arrival. This entry disadvantage diminished with years of residence in New Zealand. However, there was substantial diversity in relative labour market outcomes. While immigrants from English speaking countries had relatively small initial differentials that tended to disappear within 10 to 20 years of residence, Asian and Pacific Island immigrants had larger initial differentials that were increasing over the study period, and, in some cases, these immigrants were predicted not to reach parity with natives over their working career.

These results suggest that while net migration has added numbers to the New Zealand population over the longer term, the incoming migrants may, in fact, not be a complete replacement for citizens who departed (at least in the short run), despite being apparently higher skilled on average.

Policy changes to respond to these results could target inflows, outflows or both. Selection and settlement policies focus on finding the immigrants most likely to settle well, and helping them to integrate quickly into New Zealand. Finding ways to tap the New Zealand diaspora is one way to cope with outflows. Policy questions relevant for both inflows and outflows include trans-Tasman issues and the ultimate objective of increasing the relative attractiveness of New Zealand. Each of these is discussed in turn. What is obvious from the start is that no one policy setting can affect all aspects of the brain drain issue.

Evidence points to a brain exchange (or a replacement migration), rather than a brain drain. In addition, evidence on the labour market outcomes of immigrants indicates that the degree of substitutability of immigrants for people departing from New Zealand is questionable. These points suggest that New Zealand could gain by focusing on settlement issues and on better helping immigrants integrate into the labour market.
Settlement policy is part of a broader immigration policy framework. But immigration policy directly affects only the arrivals of those who are not New Zealanders, Australians or from the few Pacific Islands with which New Zealand has special arrangements. If it were considered that outward flows of people are a problem in and of themselves, then an increase in the inward migration target could be one solution. However, it is unclear whether it is sustainable to increase the level of inward migration much further, given that it is already running at about one per cent of total population.

It is also worth noting that patterns of movement are changing and this can influence how immigration and settlement policies should be designed. It is often assumed that international migration is all about people intending to settle permanently in a country, or at least intending to stay for a very long time. The terms ‘permanent and long-term migrants’ and ‘settlers’ are thus commonly used.

However, as globalisation trends continue, evidence suggests that migration is less likely to be permanent and long term (Bedford, 2001). People come and go, and travel more frequently than they used to. While the two terms above may have been suitable earlier in the 20th century, reality has changed at least in the last 2 or 3 decades. In fact, only a small part of the total movement of people in and out of New Zealand is this classical ‘settlers’ migration (Bedford, 2001). Therefore, there is an increasing need to give greater attention to the so-called ‘short-term’ flows in public and policy debates. For example, work permit policy may need to continue to develop a role as important as residence policy in the immigration framework. Permanent and long-term migration statistics will be less and less useful, while data on temporary movements will become more and more important.

Policy makers might well be concerned if there were significant, ongoing net outflows of the high-skilled. But even in this case, immigration and settlement policies could only help marginally, as much of the inflow, and all of the outflows, are not directly influenced by government policy.

Some have argued that rather than focussing solely on the apparent brain drain, New Zealand should endeavour to attract expatriates (see for example, Bedford, 2001). The means to implement this strategy are unclear. Meyer and Brown (1999) identified 41 expatriate knowledge networks that have been identified around the world. Their list only includes networks that have an explicit purpose of connecting the expatriates among themselves and with the country of origin and of promoting the exchange of skills and knowledge. These networks differ in size, scope, objectives, activities and structure, and are tied to 30 different countries, such as Ireland, France, Korea, Taiwan, and Norway. New Zealand did not have an explicit network identified by the authors.

As pointed out by Meyer and Brown (1999), it is often difficult to assess the success of these networks in terms of their contribution to the development of the source country. The types of exchanges that take place between network members and the national community — for instance, email and data exchanges, virtual chat rooms, informal advisory opinions etc. — do not always bring tangible, visible or immediate results and may not allow for a statistical assessment. However, it does
not mean that these exchanges are not important (see Spoonley, Bedford and Macpherson, 2000).

It may be that some way of harnessing the potential of New Zealanders overseas can be found. While the possibility is favoured by some commentators (Bedford, Ho and Lidgard, 2000; ITAG, 1999), there have been few concrete suggestions for what might be done. The experience of countries such as Ireland and India that already have implemented these kinds of policies, may provide some useful lessons (see Meyer and Brown, 1999). This strategy is increasingly becoming part of the public and academic debate, as can be seen from the recent *Catching the Knowledge Wave* conference in Auckland, August 2001 (see http://www.knowledgewave.org.nz/index_home.cfm).

Over the past 30 years, the net trans-Tasman flow has been almost always toward Australia from New Zealand. The large continuing imbalance has led to concerns about two particular areas, viz: the different criteria for third country migrants, and the fiscal costs of social security payments (see Bushnell and Choy, 2001). They have been resolved in the case of welfare payments by redesigning the underlying policies. In the case of migration from third countries small differences in criteria have been tolerated so far.

The new Australia-New Zealand Social Security Agreement could well have some implications on the brain drain debate (see http://www.nz-oz.gov.au/; Birrell and Rapson, 2001). For those who would not qualify under Australian immigration criteria, the new agreement would have an impact on their decision to migrate to Australia and/or the length of stay once in Australia. Therefore, it may affect the composition of workforce going over to Australia, as well as those returning to New Zealand.

There may be a small impact on the *ex ante* brain gain effect through migrants that seek to improve their career prospects but are starting from a low-base in terms of human capital (that is, those who tend to be classified as low-skilled). However, the impact is likely to be relatively minor because the *ex ante* brain gain effect applies particularly to those seeking higher education.

Even without convincing evidence of a brain drain, New Zealand faces a challenge. There is, and will continue to be, increasing competition in the international labour market for high-skilled people, as a result of shortages of labour in particular occupations, and the ageing of population globally.

In terms of retaining people within New Zealand, there are very few possible controls on departures of any group. The immigration rules of other countries determine their emigration options. To the extent that departures are caused by push factors (for example, there is wide speculation that the student loan rules are a driver), evening out enforcement across countries (tax agreements with Australia and the United Kingdom may be of value) will help reduce any distortions on departures (For more information on the student loan scheme in New Zealand, see the annual report to June 2001 prepared by the Ministry of Education, Inland Revenue and Work and Income New Zealand). In terms of attracting high-skilled workers, there may be scope for improving settlement and selection policies as mentioned earlier.
Ultimately, New Zealand should focus on increasing its attractiveness (for attracting highly-skilled immigrants and return migrants, and retaining existing residents) by improving the economic position of New Zealand relative to competitor countries. One way is to raise productivity more generally, leading to higher wages to attract the high-skilled. Another way is to encourage the use of more high-skill technologies, so that the high-skilled have opportunities here. However, these two strategies are endogenous with attracting and retaining high-skilled people. Therefore, both need to be part of a longer term strategy; there is no quick short-term solution.

Summary and Concluding Remarks

From an economic perspective focused on well-being, there could be a cause for public policy concern if any of the following held in the medium to long term:

- There were net outflows of the high-skilled, and no rationing of entry to the destination country. Without rationing, there is no *ex ante* brain gain effect (where people are generally encouraged to upskill by the prospect of high returns from emigrating), and only a negative effect on the human capital of the source country from emigration.
- There were net outflows of the highly-experienced. Age is an indicator of the depth and breadth of work experience. Thus if New Zealand were losing those in the middle labour force age groups, this could be cause for concern.
- There were net outflows of the high-skilled and limited return migration. Net outflows of the high-skilled can be useful if these people come home still higher-skilled at some point in the future. If they do not, then this may be a problem.
- There were other significant negative social, fiscal or economic impacts from net outflows of the high-skilled.

The key findings from the data are:

- *There are no net outflows in the medium to long-term* — We conclude that New Zealand is experiencing a brain exchange with the world (with more people coming in than going out, and those coming in more skilled than those leaving), and a same drain with Australia (where there are more leaving than coming, but the departures are not biased towards the high-skilled).
- *We switch young adults for slightly older adults* — On age profile, we are likely to be importing migrants who are more experienced in the labour market than the emigrants we are losing.
- *There is substantial return migration* — Since the early 1980s, more than 20,000 New Zealand citizens have been coming back to the country each year.
In addition there is some evidence that immigrants often have a hard time finding jobs to make use of their high skills. If immigrants are not good replacements for those leaving New Zealand, then this could lead to loss of human capital. Enhancements to immigration selection policies (to identify those likely to settle best) and to settlement policy (to the extent that assistance to immigrants after they arrive can help) may reduce concerns on this point.

There is a large population of overseas-resident New Zealanders. While many people consider that some kind of ‘remote mobilisation’ of these people is a good idea, there are but few concrete suggestions for how to go about it.

Policy makers should monitor the skill distribution of immigrants and emigrants. Changes to improve the quality of the data available from arrival and departure cards might pay dividends in the longer term.

Even without compelling evidence of a brain drain, it is sobering that so many New Zealanders continue to choose to leave New Zealand over the long term. There is, and will continue to be, increasing competition in the international labour market for high-skilled people, as a result of shortages of labour in particular occupations, and the ageing of population globally. Ultimately, New Zealand should focus on increasing its attractiveness (for attracting highly-skilled immigrants and return migrants, and retaining existing residents) by improving its economic position relative to competitor countries. This is something that cannot be done overnight.

To conclude, there is a need to move the public debate on immigration and emigration away from reactions to short-term and volatile numbers, to a longer-term perspective, and towards a more sophisticated conception of what will be an ongoing trend — that is, the increasingly free flow of people, including New Zealanders, around the globe.

References


The authors would like to thank particularly Richard Bedford, David Maré, and Grant Scobie for valuable input and comments on numerous previous drafts and antecedent papers. Helpful comments were received from two anonymous referees. All errors, omissions and views in this paper are those of the authors.
REVIEW

Corporate Social Responsibility


Reviewed by Stephen King

To a naïve economist, like myself, an introduction to corporate social responsibility (CSR) is similar to a first reading of Alice in Wonderland. It is slightly amusing, full of charm and yet clearly very silly. In some ways CSR sounds innocuous; a concept that ‘encapsulates the notion that companies have obligations that extend beyond shareholders’ (Cronin, 2001:1). But it is based on farcical ideas that would make the Mad Hatter blush. For example, a foundation principle of CSR is that business firms ‘control the financial and economic aspects of society and are possessed of monetary power far greater than the world’s governments’ (French, 1984: cited in Cronin, 2001).

CSR is a mish-mash of (sometimes contradictory) ideas. Some elements of CSR would receive general agreement — for example the idea that it should be illegal for firms to falsely represent their products. Other elements - such as the idea that firms should ignore local labour markets and differing stages of development in their hiring practices across countries - may have superficial appeal but are shown to be undesirable when subject to rigorous scrutiny.

CSR, at least in Australia, seems to boil down to little more than a call for business to donate more to charity (Duncan, 2001). This appears reasonable, although the memory that HIH was ‘an extremely generous benefactor to a wide range of worthy causes’ might lead to further thought and reflection (Brunton, 2001). However, the effects of CSR reach far further, by stating as unarguable that firms have an ‘obligation’ to communities that goes beyond the idea of mutually desirable trade and, potentially, beyond the market place. For example, banks in Australia have been castigated and threatened with government intervention for closing unprofitable rural branches. Vilification of the banks is based on the idea that the banks ‘owe’ rural communities a service, even if this takes money away from shareholders (Jones, 2001). This idea appears to be based on a wrong-headed notion that companies gain through trade but customers do not, so that companies owe their customers a return beyond profit-maximising behaviour.

Given the growth of importance of CSR and its apparent widespread acceptance, David Henderson’s short book is a timely reminder that much of CSR is both silly and dangerous. He provides an excellent critique of the arguments
propounded by CSR advocates and paints a (rather terrifying) picture of the role CSR is playing in modern corporate life.

Henderson provides a structure for analysing CSR and distinguishes between two types of related corporate behaviour. First, firms might adhere to various CSR principles because this helps to raise profits. A good corporate image can be good for business and taking account of workers, suppliers and customers almost always makes good business sense. This is nothing new. Alternatively, firms might engage in activities that lower profit under CSR even though they face no legal obligation. This response to CSR accepts the view that profit-maximising activities are not the best way for a firm to contribute to society. As Henderson states, ‘the idea of maximising profits is replaced by that of achieving ‘acceptable returns’, while business’s ‘contribution to society’ is explicitly — and wrongly — viewed as distinct from its profit-maximising activities’ (p. 41).

It is through this second type of response that CSR can undermine the welfare of society. As Henderson notes, the response reflects a failure ‘to understand the rationale of a profit-based market economy’ (p. 63). In particular, it fails to recognise the role of profits as an incentive for firms and the role of mutually voluntary trade in making both buyers and sellers better off. For example, one of the key issues for CSR involves wage payments and labour laws in less developed countries. Large multinational companies based in the developed world invest in these countries and employ workers at wages and conditions that would be illegal in their home country. Such practices are abhorrent to CSR adherents and should be eliminated by the relevant companies. Under CSR, multinational companies have an obligation to support ‘humane’ labour practices.

But before concluding that the hiring practices of multinational companies are immoral and a source of harm to the poor, consider the effect of both the policies and their elimination. Does the hiring of poor impoverished workers in the less developed countries make those workers worse off? No! As Henderson notes ‘the adults who voluntarily seek employment with foreign-connected firms, on terms that they are aware of, do so in the hope and expectation of becoming less impoverished’ (p. 70). While wages and conditions of those workers might appear unacceptable to residents of the first world, they are more than acceptable to those in poor, overpopulated countries where the alternative would be even lower pay in domestic firms or greater impoverishment and starvation. But surely it would be better if those same workers received even higher pay? Not necessarily! The abundant supply of cheap labour creates the incentive for multinational firms to invest in unstable countries ruled by corrupt governments because it is profitable, despite the political risks. If well-meaning CSR advocates in developed countries forced the wages of the local workers up, foreign investment in less developed countries would dry up. It is cold comfort to a starving peasant that he or she would receive a high wage if only they had a non-existent job (Krugman, 2001).

If the problem, at least partially, is corrupt government and political instability in less developed countries, should business have a social responsibility to pressure reform in those countries? Henderson argues that such a stance is fraught with danger. Why should business be the judge of acceptable behaviour?
This is the role of government. Implicitly, the CSR position elevates business to a role in determining socially desirable policies.

The CSR process involves forcing business to lead government. By sacrificing profits to attain an inflexible CSR position (for example, see Henderson, page 65, for the principles of ‘eco-efficiency’), firms harm their customers — through higher costs and prices — and make themselves less competitive. This in turn harms their suppliers, their workers and their shareholders. As these ‘socially responsible’ firms lose market share, capital and labour to their ‘less responsible’ competitors, they will be forced to either abandon their CSR position or get their rivals to adopt it. The latter will involve corporations who adopt CSR lobbying for government intervention on their competitive rivals. Henderson captures this nicely (p. 79) with a conversation between two executives of rival mining companies: ‘My organization is run by Greenpeace today, and it is my job to ensure that Greenpeace is running yours tomorrow’.

While Henderson provides an excellent overview of CSR and effectively critiques many aspects of CSR, I have two reservations about his book. The first relates to style. In my opinion, the book takes a long time to ‘get going’. The first thirty pages (of a ninety-four page book) did little to convince me that CSR really was a problem. In these pages, Henderson defines and categorises CSR and discusses its emergence and growth. But these first three chapters do not create a sense of the importance of Henderson’s topic. It is not until Henderson starts to quote business and political leaders, particularly from page 29 on, that the true extent of CSR and its potential dangers become clear.

My second reservation is about the nature of Henderson’s critique. In many ways, I think that Henderson gives CSR more than its due. While Henderson mentions the issues of mutually beneficial trade and the nature of firms as voluntary groupings of disparate individuals, he does not emphasise that these issues are the key factors in understanding why CSR is flawed. Firms only make abnormal profits by helping people in extraordinary ways through the market place. And those very profits, in the absence of government intervention, contain the seeds of their own destruction, as they lure in new firms who drive down prices, raise service and compete away the abnormal profits. This is how firms benefit society. There is not a conflict between profits and social welfare. Rather it is the lure of profit, in a free market, that creates social welfare.

Further, before calling for corporate social responsibility, it is necessary to ask what a corporation actually is. A firm, no matter how large, is simply a group of individuals who voluntarily join together through contractual arrangements. These individuals are part of society. To talk of corporate responsibility to society as separate from the corporation as part of society is nonsense.

By considering the corporation as part of society, the true nature of CSR becomes clear. CSR policies fall into two groupings. First, many CSR policies simply aim at redistributing society’s wealth from those who act within corporations (and particularly those who participate in corporations by providing funds as shareholders) to those who do not participate in corporations. Such
redistribution might be worthwhile. There is no reason why a market-based economy will lead to an equitable division of wealth. However, such redistribution policy is up to government, not company managers, to decide.

Second, CSR policies often involve using corporations to impose on society particular principles that are not endorsed by most of the population. If most people supported CSR policies then firms would find it profitable to adopt those policies. Consumers would be willing to pay more for products from a firm that does not ‘exploit’ cheap labour or that adopts certain environmental practices. Similarly investors would be willing to forgo some return to fund CSR compliant firms and employees would be willing to receive lower wages to support those firms. That CSR policy must be imposed through high-level lobbying and government pressure reflects that most people in society do not support those policies. When faced by a ‘responsible’ product at a high price and a cheaper ‘irresponsible’ product, most people prefer the cheaper product. CSR is an attempt by a minority who support certain policies to use corporations to impose those policies on the majority of the population who do not accept them.

Henderson has written an important, timely book. However, I see the risk of CSR differently to Henderson. Over time, CSR is unlikely to survive through voluntary corporate compliance that reduces firm’s profits. CSR thrived in the economically buoyant 1990s. In tighter economic times, CSR will lose much of its gloss and senior executives that support it are likely to be replaced by hard-nosed realists. The real long-term risk of CSR will be misguided government policy. If CSR leads governments to pass laws where the social benefit is outweighed by the cost to business and the community, then CSR will create significant loss.

References


Stephen King is Professor of Economics at the Melbourne Business School.
**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)

### Dumbing Down: Some Thoughts on a Phrase of our Time

**Don Aitkin**

Dumbing down has a fine ring to it, the result of percussive alliteration reinforced with energetic direction. The term has interested me for some time, partly because of its novelty, partly because I wondered what had brought it forth, and partly because intuitively I rejected what I thought was its message. What follows is an essay on these themes, with a commentary on what seems to me to have happened in Australian education over the past fifty years. I accept from the beginning that the debate about standards (for that is what ‘dumbing down’ is about) is a rich one. This essay is intended as a contribution to that debate, not as a kind of closure.

It is plain at once that ‘dumbing down’ is a critical term: someone is doing something bad to someone else. What is happening, and who are the actors? While my first encounter with the phrase a year or two ago was Australian, the term has an American ring. The Internet is a great source for American catch phrases, but my search engine produced a truly embarrassing richness: more than 32,000 references. The first hundred of them, however, made clear what further search was only to confirm. The phrase is amazingly popular, but the stock of arguments is quite small. The positions and arguments are inter-connected, and all of them possess sadness, nostalgia and a kind of bitterness or anger. I set them out in the order of descending generality.

**How the Term Dumbing Down is Used**

The first is a strong critique of contemporary Western civilisation, whose people live in a ‘moronic inferno’ characterised by trivia, sensation-seeking and

*Don Aitkin is the recently retired Vice-Chancellor of the University of Canberra.*
drugs. Problems are too hard, and people have retreated from trying to solve them into a kind of numbness. The system can be described as a ‘dumbocracy’, the rule of cleverness without wisdom, and its evils can be seen in all walks of life (Mosley, 2000 is an example).

The second is a version of this position that excoriates the mass media and finds them wanting and therefore responsible. According to critics in this domain, television, radio and the newspapers have abandoned serious news and challenging artistic programs for a safe, sugary ‘infotainment’ world that does not offend advertisers and keeps viewers/listeners/readers in a mentally dormant condition. Movies are similarly dumb, because they’re made for dumb moviegoers who get restless if they are asked to see anything challenging. Museums are becoming dumb because the interactive and digital accompaniments designed to attract people, it is alleged, actually inhibit thinking.

The third provides a possible ‘design’ of this state of affairs. Charlotte Iserbyt’s (1999) The Deliberate Dumbing Down of America points the way. Here the plotters are ‘reformists’ and ‘do-gooders’ who have interfered with the American education system, replacing rigour with warm, fuzzy and mushy stuff. She is not alone. Similar critiques abound, though not all the plotters are seen to be on the left. There is a strain of anti-capitalist, anti-corporation sentiment in all this, too, the argument being that corporations simply want a nation of docile consumers, and the deplored changes to the education system are intended to achieve such an outcome. Another group thinks that it is governments who are responsible, because they want to regulate everyone, and are therefore interested in people not being able to make decisions for themselves.

The fourth provides particular contexts for such a general attack. Teachers don’t have ‘real’ learning to acquire any more, but are fed courses on ‘diversity’. Widening access (to schools, colleges, universities, professions) must lead to a decline in quality. Sermons in church are now soft and mushy, and congregations are not forced to think and reflect. Fewer Americans are able to converse in other languages, or to count, or to undertake physics, or to do whatever the critic thinks is important, than was once the case, than would be ideal, etc. More people believe in the healing power of crystals, pyramids, copper bands, and so on and fewer in real medicine (or in God’s word).

The fifth is a specific attack on what are seen as falling standards, the cause or root of the changes already referred to. Any example of a new standard is likely to be attacked as a ‘fallen’ standard, and to be an example of dumbing down. In a splendid British example, a slight increase in the proportion of A-level results in high school exit examinations was seen as a disaster by critics (an obvious sign that standards had slipped) and as a great success by defenders (an obvious sign that standards were rising).

The sixth is about the decline of skill, and is reminiscent of Harry Braverman’s (1974) lament in Labour and Monopoly Capital: The Degradation of Work in the 20th Century. What makes this theme most interesting is that the skill being highlighted is computing. There are continuing debates about the cost in skill utilisation of so-called ‘user friendly’ personal computers, continued
exposure to which is said to lead to a loss of programming skills. How dumb does a web-site have to be to attract sufficient hits to make it worthwhile, asks a complainant in an adjacent field of endeavour. No one wants to have to do hard work any more, a lament echoed elsewhere in these conversations.

Australian References

Australian references appear from time to time, and by adding ‘Australia’ as a descriptor you can find them all together. By doing so you lose much of the paranoid flavour, and of course the plainly American material. There’s nothing about UFOs, either. (For those who don’t immediately see the connection, it is claimed that governments have dumbed down the population so that people don’t ask about unidentified flying objects, preferring to believe government reassurances that there aren’t aliens, UFOs and the like.) A striking characteristic of the Australian material is the frequency with which ‘dumbing down’ appears in the sub-head or title of the article or story, even though it was not a major element in the text. My guess is that sub-editors find it a catchy phrase to use, and employ it even when the main message is about something else.

An Australian example is ‘The Dumbing Down of Australian Professions’, by Padraic P. McGuinness, which appeared in the Sydney Morning Herald on 15 April 2000. It begins ‘Why are our teachers so stupid? Why are our doctors so silly?’ These opening questions are not followed, as you might expect, by examples of stupidity or silliness; that is simply taken for granted. The explanation offered by McGuinness is the democratisation of tertiary education started by R. G. Menzies in the 1950s, which produced the first great wave of expansion in higher education, and the indifference of the Whitlam Government to quality control, which took the brakes off thereafter. Once upon a time teaching ‘creamed off the best of the working-class and poor school students, and for some time we had what was probably one of the best teaching services in the world’. Now what we have are ‘tenured mediocrities and failures’ in our schools and universities, and ‘a medical profession full of academically intelligent but ill-educated, ill-read and bored GPs’.

I would agree that this is not Paddy McGuiness at his best, but the ‘dumbing down’ of Australia is a steady theme of his, in Quadrant, of which he is the current editor, as well as in his newspaper essays. The article exemplifies a good deal of the material that appears under the heading of ‘dumbing down’. To begin with, there is no evidence or example of the failings being criticised. It is as though it is all so obvious that there is no need to do anything other than utter the magic words. (In fact, McGuiness does not use the phrase ‘dumbing down’ in the body of his text; its use in the title may well be the work of a sub-editor.) In none of the articles that I have read has there been any worked-out account of the process through which a society is supposed to have become dumbed down. The argument is rarely more than ‘Once that, now this. Woe!’ And of course the decline is usually somebody’s fault.
Next is a firm assumption, or set of assumptions, about the way the world really is or ought to be. Such assumptions usually have to be inferred, because they too are not spelled out. In McGuinness’s case what we see is the old IQ assertion: only a small proportion of the community is intelligent enough to benefit from high school or university education. Any expansion of the system must reduce quality, because there aren’t enough ‘bright’ people to do the work that only such bright people can do.

A partner of this assumption is the view that life is tough, and that education systems should reflect this reality. Dumbing Down Our Kids by Charles Sykes (1996) contains a set of rules for life that have been widely circulated by fax and email. Some of them are:

Rule 1 Life is not fair; get used to it.
Rule 2 The world won’t care about your self-esteem. The world will expect you to accomplish something before you feel good about yourself…
Rule 4 If you think your teacher is tough, wait till you get a boss. He doesn’t have tenure…
Rule 8 Your school may have done away with winners and losers but life has not. In some schools they have abolished failing grades, they’ll give you as many times as you want to get the right answer. This, of course, bears not the slightest resemblance to anything in real life.

The McGuinness essay also lacks much sense of time or any sense of scale. Yes, he does refer to a past and describes it. To be fair to him, the great majority of the articles I’ve read did not — it was assumed that we all knew about that. It is not obvious from the article, however, that McGuinness thinks it important that the Australia of 1947 contained 7.5 million people, that of 1982 had 15 million, and that of today has passed 19 million. Wouldn’t these increases have had an effect? Was no expansion of education necessary? It’s as though societies are static, or frozen.

I don’t really need to defend Australia (or the USA or Britain) from the charge that it is dumber than it used to be, and/or that some named or nameless persons have accomplished such an outcome or are endeavouring to do so. A few minutes’ thought, and some comparison with the past, will make such a proposition implausible if not preposterous. The Guardian (London) devoted three issues to the proposition in November 2000 and came to the conclusion that it was simply rubbish. The reverse was true — on the evidence society was ‘clevering up’, and if this was a problem to ‘the angry old men of the dumbing down debate’ (Madeleine Bunting, Guardian, 13 November 2000), they would just have to get used to it. I would argue that the Australian evidence points in the same direction, and could amass a stack of data. Alas, the data are usually dismissed by critics as irrelevant. A great deal does depend on what we mean.
Understand What Is Being Assailed

So it does seem to me to be worth trying to understand what is being assailed, and why, and what can be said about it. I can remember taking my youngest child to a parent-teacher night for kindergarten in 1987 and discovering myself in a fight between tough-minded and tender-minded parents about what school was really for. These are matters about which people do feel strongly, and about which Ministers inevitably have to form views. Not only that, people who have invested a lot of time and effort in something expected to get a return for it: if you are a programmer, for example, user-friendly machines may well make you feel that your livelihood is imperilled.

Let me set out the educational expansion of the last fifty years or so in Australia as factually as I can, and then reflect what meanings may be derived from it all. At the end of the second world war most Australian children did not complete high school, and only a very small proportion (2 per cent of my age group in 1954) went to university. High schools were relatively scarce, and staffed for the most part by teachers who had university degrees in the areas they professed. Those schools were also highly disciplined, teachers had authority and status within their communities, not simply in the classroom, and pupils wore uniforms prescribed by the school, whether they were in public, Catholic or Protestant schools.

Today the university system is twenty times larger than it was; there is a high school in almost every Australian town, a little more than 70 per cent of children complete high school, which is not as disciplined an environment as it once was, and most Year 12 students expect to go to university. Three engines drove the expansion of tertiary education. The first was the need to increase skill levels in the workforce. The second was political pressure from parents wishing their children to have a more enjoyable working life than the one they knew (and reasoning that education was the necessary escalator). Third, and not least, was the outsourcing of professional training by the professions and the highly skilled white-collar occupations from the workplace to the university. Human knowledge (that is to say, what academics define as ‘knowledge’) has multiplied around fifty times since the end of the second world war. Specialism abounds. There is a continual cry for a new breed of generalists and synthesisers, but it is vastly more difficult to be a generalist in the early 21st century than it was in the early 20th. There is just so much more to know and more to synthesise.

These broad changes have led to others at the more personal level. For example, today’s children can expect to spend most of the first thirty years of their lives in serious education. There is just so much more to learn. Most of the occupations that people have today did not exist in 1947, and most of the occupations people had in 1947 no longer exist or have been radically transformed. The school curriculum is actually rather more conservative than the workplace, but it too has undergone great change. Today’s young women will on average have their first babies at 27, and will have, again on average, only 1.7 of them. The generation of 1947 was more fecund and started baby production
earlier. Divorce was difficult and uncommon in 1947; it is neither today. The churches were much more powerful in 1947, there was much more of a single morality, at least in public, and Australia was a much more solidary, ethnically homogenous society. Not everyone liked this latter state of affairs. Some left Australia altogether because of what they saw as intolerance and an excessive need for conformity. Others became rebels. Others worked hard to push society’s norms into a more progressive stance.

It should not be surprising that some people find not to their taste the changes from then to now (and the counterpart changes in the USA and Britain). It should be no less surprising that some people will find today’s society and its context more threatening than the earlier society that they remember. I would generally agree with those who argue that today’s Australia lacks a sense of purpose or mission comparable with that of the 1950s and 1960s, and I think that missing element does affect the attitudes of many young people to life. The urge to condemn, and to construct a ‘decline and fall’ picture, will for some people be a strong one. I could construct one myself if I let go of my historian’s feeling for balance. I could, for example, complain that no one these days seems to know how to use the apostrophe properly, whereas in my day we were taught the parsing and analysis of English sentences and exactly when and where to use the apostrophe. What on earth has happened to language, I could thunder. What do the schools think they’re doing? And so on.

But balance reminds me that only the A classes gave much attention to the niceties of English grammar and syntax, and that we (I include Paddy McGuiness here, much my age and no less well educated) made up a small proportion of a small proportion. I need to remind myself that around half of the Australian people today were either born overseas or are the children of parents, one of whom at least was born overseas. English is a second language for many of them, and the English language is itself evolving. I would be prepared to wager that most people in my days didn’t know where to put the apostrophe either, but they did not have to use written English much in their jobs, so they weren’t caught out. Maybe the apostrophe is on the way out. The Germans use a final ‘s’ as a possessive, as we do, but they don’t use an apostrophe to separate it from the noun. Whatever else has happened in the last fifty years, there is no doubt that much greater proportions of Australians communicate within our society than was once the case, in text as well as in pictures.

If I return to the themes at the beginning of this essay, I have to say that I thought that the expansion of education would have had a much more powerful beneficial effect than it appears to have had in the areas of radio and television. Parents today worry about the evil effects of television; mine were adamant that comics would rot our minds. On the bright side, newspapers are a lot better than they were in the 1950s, and having around 3 million university-educated people in our population has raised the general levels of discussion, argument and confidence of public debate. If you don’t think so, I invite you to read the Hansards of the early 1950s or the newspapers of that time. Nonetheless, there is still a good deal of the ‘bread and circuses’ mentality in our culture. I wish there
wasn’t as much as there is, but I balance against it the wonderful advances there are in music, the creative arts, theatre and literature.

Like the critics, there are times when I want to argue that it is time that some people accepted that hard work generally precedes lasting success and achievement. When I say such things I have to avert my eyes from the people who make lots of money by pushing other money around, or by owning companies that exploit a monopoly situation of one kind or another. There are times when I want the acceptance of responsibility rather than the demand for rights. And there are times when I wonder whose fault it is that Australia is not the way I would like it to be, and who allowed it to get this way. I don’t go far down this track because I can see that it is the generation of my parents and the generation of which I am a part (and indeed Paddy McGuinness is a part), to which much of the blame has to be levelled for whatever shortcomings are seen.

So I look at the brighter side, the sheer exuberance, curiosity, tolerance, creativity and hard work of contemporary Australia, and reflect that maybe we didn’t do so bad a job, all things considered. Because I work in the education business, I know quite a lot about changes in curricula, the entry into the university of new professions, like medicine in the 19th century, and education in the early 20th, and law finally becoming respectable by sidelining the admission boards, and accounting making it in, and then nursing, and computing and tourism and public relations; even journalism. There’ll be new professions as the new century advances, and people will sneer about them too. The notion that some areas of knowledge are not really suited to university study is much affected by the location of the person making the judgement, as I have observed over almost half a century of work in higher education. A good deal of the clamour, I think, is about status.

‘Twas Ever Thus

There are those who argue that today’s nurses don’t know how to fluff a pillow, or that today’s journalism graduates don’t know how to make it in the real world, or that today’s doctors are over-trained, or that those who can do and those who can’t teach. ‘Twas ever thus. One of my tasks for the Leaving Certificate in 1953 was to work my way through Cicero’s De Senectute, a reflection on old age written by Cicero in his last years. I don’t recall the Latin now, and couldn’t translate it, but Cicero knew a thing or two. One wise observation of his was that life was always better some time ago, and another that sooner or later people told you as great discoveries things that you had known when you were young. Reading about dumbing down reminded me that I did get something of substance from all those long periods in Latin.

And that takes me to standards. I’m actually glad that I learned Latin, because I became a writer, and Latin (along with the French and German I learned at the same time) has been very useful to me as a writer of English. But I couldn’t suggest that because it was good for me it should be compulsory for everyone else. It once was, of course. In the 19th century it was the basis of entrance into
university. Once any writer of quality would quote from the Bible and expect his readers to recognise the reference. At another time writers would interrupt their flow with a bow to ‘the immortal Bard/the poet/the blind singer’ and put down a couplet. Again, they knew their readers would know whom they meant and why it was relevant. Do that today and no one much would know what you were talking about. That is not, in my view, because standards have slipped, but because each generation has its own canons, icons and reference points, and their status changes all the time.

University campuses echo with the laments of falling standards. I can remember Henry Rosovksy telling us in the late 1970s that the proportion of A grades in Harvard College had doubled since 1945. The inference was that standards had slipped. But maybe the students of the 1970s worked harder. I’d be prepared to bet that the students of today do. In any case, Howard Gardner’s powerful Frames of Mind (1983 and many later editions) gave me the neurophysiological and educational explanation of why it was that all my life I had seen the numbers of people going to university increase and labour there successfully. The truth is that we are all intelligent, across a number of dimensions (Gardner says 8½), and it is love, encouragement, motivation and preparation that distinguish us in the ways we benefit from our native gifts. Those who see dumbing down around them need to read Gardner, but it is a fair bet that they haven’t and won’t.

Entering students aren’t prepared as well as they used, to be in mathematics, or physics, or English, or history, or whatever, comes another complaint. Built into this lament is the assumption that schools are there to prepare students for university. I used to think so too, but now I see that outcome as just one of the tasks schools are expected to undertake. My father, a maths teacher in high school for much of his working life, once pointed out to me that there were just so many hours in the school day, and if I wanted something inserted into the curriculum I would need to suggest what should be dropped. Some of the critics of dumbing down have a ready answer: ‘Concentrate on the three Rs!’ they will cry. Every now and then, looking at yet another poorly spelled sign or an advertisement with an ‘it’s’ horribly rendered, I sympathise. But we’re long past that.

The schools have to fit into the curriculum a lot that wasn’t there in my day. I went to good high schools and was well taught in them, but I learned no Australian history and no social science of any kind, my history stopped at 1900, computing skills were unheard of, we did virtually no oral work at all except in languages, we knew nothing about the world, and learned nothing at all about sex save behind the dummies (and that was in parts ludicrously incorrect). Preparation for work? You’re joking. We were streamed at an early age, and the choices were final and exclusive. Once you were on the science side or the arts side that was that, and streaming started at about age 12. Today’s schools have an unenviable job, because apart from everything else they have become the dumping place for the education that children used to get (or not get) at home but for which today’s busy parents do not have time. But in 2002 we have a lot more choice. Not only do kids have much more choice than was once the case, but so do their parents.
There is a much greater variety of schooling contexts available today than there ever used to be. If you want it, it’s there somewhere.

To insist that there are defined standards and that everyone must abide by them is a really fine piece of arrogance, one that assumes the standards preferred by the speaker must be those chosen by any right-thinking person. In fact, the diversity of our society and the increased levels of confidence possessed by its citizens mean that there is considerable diversity in the standards preferred by them. I would agree that if there is too much diversity then it may be hard to see what holds our society together other than propinquity. But a single set of standards is a cast back to the kind of totalitarian regimes that the great majority of those opposed to dumbing down would reject at once. I’m afraid that diversity rules, OK?

Finally, there is the worry about de-skilling, the dumbing down of the workforce, and the loss of knowledge. Braverman saw the Taylorist mode of industrial production as having denuded craft workers of their skills and created an unskilled working class that was alienated and without pride. But in industry after industry, most notably the automobile industry, robots are replacing the workers who knew only how many wheel nuts to screw on and how quickly to perform that task. Today’s workforce is all about skill, and the investment in workers is many times what it used to be. One consequence is that there are fewer workers in industry, another that hardly won skills do not have the shelf life they once had. The flight engineer who monitored the four propeller-driving engines of the DC-6 and the Constellation had only a fifteen-year career. Jet engines, simple and reliable, replaced the old radials, and they required no engineers or monitoring. The shift from the large mainframe to the desktop PC displaced thousands of skilled workers, and the shift from the PC to the pocket-sized Palm and its counterparts will displace thousands more. My guess is that any serious study of skills and deskilling over time would show that the process of de-skilling and re-skilling is one that all human societies have known. There are more or less humane ways of controlling the process, but the notion that we can somehow prevent it seems almost ludicrous to me.

There seems to be no end to the technological changes to which we are subject, and all of them mean that we have to keep on refining our knowledge and learning new skills. To do so all the time is tiring and frustrating, but it can also be stimulating and a continued affirmation of life. Of course, there is an alternative. We can sit on our little pile of accumulated knowledge, pretend that it’s everything, and cry ‘dumbing down!’ about the others when we see ourselves being passed by. That seems pretty sad to me.

References


The author gratefully acknowledges the benefit of the comments of two anonymous referees.