EQUITY

AS A SOCIAL GOAL

Cathy Buchanan and Peter Hartley

New Zealand Business Roundtable
March 2000

First published in 2000 by
New Zealand Business Roundtable,
PO Box 10–147, The Terrace, Wellington, New Zealand
http://www.nzbr.org.nz

ISBN 1–877148–61–X

Production by *Daphne Brasell Associates Ltd*
Printed by *Astra Print Ltd*
Cover photograph by *Julia Brooke-White*

Contents

List of Figures

List of Tables

Acknowledgements

We thank Roger Kerr for his support throughout this project. We are also extremely grateful to Richard Epstein for providing extensive, and very valuable, comments. We also received valuable comments and advice from Joe Barnes, Peter Bushnell, Greg Dwyer, Michael James, Ross Parish, Larry Temkin and Bryce Wilkinson.

We have not necessarily incorporated all of the suggestions of the referees into the book, however, and they cannot be held responsible for any of the errors or interpretations that remain.

Our discussions of discrimination and native land title in Chapter 7 draw on Chapter 8 of Hartley (1997) *Conservation Strategies for New Zealand*, which was written by Michael Warby.

1
INTRODUCTION

The efficient allocation of resources is the focus of many analyses of public policy. While economists are often criticised for assuming that efficient resource allocation is the only goal of society, most argue that promoting efficiency is just one of the foremost ends of social interaction. Efficiency losses are relevant to determining the advisability of any activity, even when other objectives are pursued.

Economists have been prominent among those who have advocated liberty as a fundamental principle governing social interactions. A respect for individual autonomy and freedom to arrange mutually acceptable exchanges of goods and services might be supported on the grounds that adherence to these principles promotes efficiency. A commitment to liberty as a fundamental principle, however, goes beyond determining an attitude to economic relations (narrowly conceived) to encompass support for freedom of expression, religious freedom, democratic participation in governance and an independent and impartial legal system. In some cases, a commitment to liberty might conflict with a desire to promote efficiency.

The pursuit of efficiency also is often criticised as a goal because it ignores the distribution of the net benefits from economic activity. As usually applied, the criterion of efficiency treats a given level of net benefits the same, regardless of who enjoys the benefits or who bears the costs. Roughly, the efficiency criterion assumes that 'a dollar is a dollar' no matter who earns it.

Equity is, in part, concerned with the distribution of the net benefits from economic activity. The term 'equity' also refers to fair or equal treatment. People ought not to be victimised or suffer from discrimination, and everyone ought to have the same civic and legal rights and responsibilities. Policies suggested by the latter notion of equity may promote efficiency, so the two goals can reinforce each other. Equity, however, is often used to refer to unequal treatment intended to reduce inequalities arising from other sources. Policies ostensibly aimed at reducing inequality typically come at the expense of efficiency.

In this work we attempt to clarify the notion of equity as a goal of public policy. Among the questions we shall address are the following. How can we decide what the goals of public policy ought to be? Why should efficiency and liberty be included among our social goals? Ought equality or some other notion of equity be a social goal? How might equity best be incorporated into policy?

ETHICS AND PUBLIC POLICY

Our central concern is the status of equity as a goal of public policy. Some might argue that efficiency is not only necessary but also sufficient as a social goal, and that equity is of minor importance. For example, Epstein (1995: 30), when discussing criteria useful in choosing a legal system, states:

> [T]he maximisation of social utility becomes the objective of a sound system of legal rules. Although I have from time to time been of different minds on this proposition, I have now made peace with myself and believe that the consequentialist theories – that is, those which look to human happiness – offer the best justificatory apparatus for demarcating the scope of state power from the area of individual choice.

For Epstein, efficiency is the only relevant criterion for judging whether a legal system is desirable. More specifically, Epstein (1995: 32) argues that the purpose of a legal system is "to minimise the sum of the administrative (including error) costs and the costs associated with the creation of poor incentives for individual action".

Epstein discusses whether the simple rules he derives on the basis of the efficiency formula "should be supplemented by some additional principles of justice":

> The common conceptions of justice … are consistent with the basic theory outlined here, and indeed are required by it. Make way for Occam's Razor. If a smaller class of assumptions can be used to account for all the relevant results, why treat the intuitive sense of justice as the irreducible primitive of the system or even as an important side constraint? It is far better to regard the various conceptions of justice as deducible from this general theory of social relations with its more prosaic efforts to elaborate the systematic relationships between administrative costs and incentive effects … Ideas of unfairness are dangerous when not moored to any substantive theory. (Epstein, 1995: 319–320)

Epstein suggests that free-floating notions of unfairness can be used to support almost any position. By parity of reasoning, public support for efficient outcomes may be tenuous unless it is secured by a respect for moral intuitions that are shared across societies and times. Liberty and respect for fair or equal treatment may only be sustained by a widespread belief in their value as ends-in-themselves.

The fundamental issue is how we value the possible outcomes of our social, political and economic interactions. A judgment about the adequacy of efficiency, equity or other goals of a society is inherently normative. We cannot agree that efficiency, equity or other goals are better than their alternatives without having some way of ranking these goals.

Ethical judgments involve statements about what ought to be done. As a matter of logic, such ethical conclusions necessarily involve ethical premises. As Hume noted:

> In every system of morality, which I have hitherto met with, I have always remarked, that the author proceeds for some time in the ordinary way of reasoning, and establishes the being of a God, or makes observations concerning human affairs; when of a sudden I am surprised to find, that instead of the usual copulations of propositions, *is*, and *is not*, I meet with no proposition that is not connected with an *ought*, or an *ought not*. This change is imperceptible; but is, however, of the last consequence. For as this *ought*, or *ought not*, expresses some new relation or affirmation, it is necessary that it should be observed and explained; and at the same time that a reason should be given, for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it. But as authors do not commonly use this precaution, I shall presume to recommend it to the readers; and am persuaded, that this small attention would subvert all the vulgar systems of morality, and let us see, that the distinction of vice and virtue is not founded merely on the relations of objects, nor is perceived by reason. (Hume, 1978: 468–470, emphasis in original)

Ethical premises necessarily underlie any suggestions for the most desirable, or the least undesirable, policies to follow in particular circumstances. This is not to say, however, that only values matter for evaluating policy goals, and that evidence concerning the likely consequences of different actions, or institutional arrangements, is irrelevant. Good intentions may be desirable, but the best intentions can often lead to the worst results.

Economics, or the study of the allocation of scarce resources to competing ends, is indispensable to any discussion of desirable goals for society. At one level, economics is a purely descriptive enterprise. The aim is to describe how resources are likely to be allocated under different institutional arrangements and in different circumstances. Such an analysis is, however, relevant to ethical discussions of equity because different ideas regarding what ought to be done often rely on judgments about the likely consequences of implementing those ideas. While ethics is concerned with means as well as ends, likely end states are usually one element in evaluating different ethical rules.

Ethics is also relevant to examining the status of equity as a social goal because ethical beliefs may be one determinant of the policies that are implemented in a democracy. Politicians often use moral arguments to justify policies. Journalists and other commentators also evaluate policies on the basis of the perceived moral worth of the policy goal. More generally, most voters have neither the time nor the information, nor a strong motivation, to examine the consequences of different policies. There is a tendency to focus on motivations and goals rather than outcomes. To gain majority support, therefore, politicians must appeal to widely shared ideals or beliefs. These common values or beliefs may include what many would regard as ethical rules or standards. Politicians may also, however, appeal to other precepts that most people would not regard as ethical rules.

PROCESSES AND OUTCOMES

A final reason for discussing key ethical concepts is that many of the dilemmas which arise when we attempt to create, reform, or evaluate social institutions also confront an individual attempting to behave ethically. Since each of us is more familiar with personal decision-making processes, it is easier to explain moral dilemmas in the context of individual moral choices. We later draw an analogy between personal moral choices and evaluating the likely consequences of different policies or social institutions.

The first ethical dilemma we address is the conflict between means (or processes) and ends (or outcomes). We have alluded to a potential problem with focusing only on means. The best intentions can lead to undesirable outcomes. Conversely, ends do not always justify means.

In the political or social realm, we can concentrate on developing institutions or modes of interaction that may lead to desirable outcomes on average, or over the long term, even if we cannot be sure about the particular outcomes that will result. Alternatively, we can attempt to institute policies aimed at achieving particular results.

Ethical rules can be viewed as constraints voluntarily placed on the choices or actions regarded as viable in particular contexts. Similarly, constitutions or laws enforced by an independent legal system constrain the policies that governments can implement. Rules that limit choices would appear to be undesirable, or at best irrelevant, constraints on the options available to the decision-maker. We shall argue, however, that rules may actually allow a decision-maker to achieve more than would have been possible without such constraints. The basic idea is that the knowledge that the actions of others, and of one's future self, will also be constrained by rules alters the feasible set of actions today.

With regard to majority rule, there is an additional reason for constraining the actions of the decision-maker. Competition among political parties for a majority of votes at periodic elections cannot be relied upon to produce policies that further all and only those goals that are desirable. Unless the types of decisions that majorities are allowed to make are constrained, democratic government has the potential to do substantial harm. This is not to say that dictatorships or one party rule are better. Democracy may be merely the least bad form of government. The real issue is deciding which constraints should be placed on what the majority is allowed to decide.

EQUALITY AS A GOAL

Concern about equity is sometimes expressed as a belief that people ought, to the greatest extent possible, enjoy an equal share of the benefits from economic activity. An immediate practical difficulty, usually ignored by advocates of such a goal, is that net benefits are in practice impossible to measure. Suppose we could ensure that people had exactly the same material resources, performed exactly the same actions and enjoyed exactly the same consumption bundles. There would still be no guarantee that they would receive identical net benefits from their labours. Factors that are impossible to measure, such as different abilities, tastes and information would mean that different people evaluate the identical actions and consumption bundles in different ways.

In practice, egalitarians usually are content to advocate equalising along certain very restricted dimensions, such as equal money incomes in a given year. It is far from clear, however, that promoting equality in such restricted dimensions would actually increase equality in the net benefits from economic activity. Furthermore, even if one could ensure equal shares of net benefits from economic activity, people could still be unequal in other dimensions, such as in their ability to benefit from leisure time, that may be even less amenable to objective quantification.

Another fundamental problem with equality as a goal is that there are two different motivations for being concerned about inequality. One is compassion for our fellow human beings who are deprived of basic goods or services. The other is envy of people who have more material wealth, or greater control over some scarce but desirable resource, than we do.

We show that the commonly used measures of inequality do not distinguish between envy and compassion as reasons for being concerned about inequality. Furthermore, while compassion characterises much of the rhetoric used to support policies ostensibly aimed at reducing inequality, in practice much of the support for such policies appears to come from people envious of others perceived to be enjoying preferable circumstances. Government policies in a democracy also tend to result in more assistance being provided for middle income families than for those genuinely in need.

Some loss of efficiency may be justified for the sake of assisting people in desperate circumstances who have inadequate access to savings, loans to be repaid from future income, or to familial and private charitable services. We argue, however, that policies aimed at appeasing envy impose substantial costs while delivering nothing of value in return.

Equality of opportunity

Some commentators have suggested that equality of opportunity may be a more reasonable goal than equality of outcome. Since no two people are the same in all respects, even if everyone had the same chance to succeed in life, it is most unlikely that everyone would end up in the same circumstances. It may appear to be harder, therefore, to ensure that people end up equal than it is to ensure that everyone has the same chance of ending up in any given circumstance.

We argue, however, that it is easier to ensure equality of outcome than to ensure equality of opportunity. One would need much more information to ensure that individuals had the same opportunities available to them than to ensure that people ended up being equal along certain measurable dimensions. In fact, much of the information one would need is unavailable and probably unobtainable. While we could in principle observe the choices people make, it is much more difficult to measure the range of all possible choices they could have made.

It is sometimes suggested that equality of access to certain goods or services, such as formal education, professional health care and competent legal representation, would be sufficient to ensure equal opportunities. These claims are naive. In order to ensure equal access to formal education, for example, much more than attendance in the same classrooms would be needed.

Nevertheless, many voters seem willing to subsidise the consumption of particular goods or services regardless of whether such subsidies actually increase equality of opportunity. Compulsory transfers of money income seem to provoke greater opposition than transfers in the form of services of equivalent cost. In fact, subsidising the consumption of particular goods and services can at best benefit the recipients by as much as a money transfer of equal cost. Moreover, services financed by taxes are often supplied by monopoly public agencies. The resulting higher costs of provision and lower quality of service imply that even less is achieved for the recipients per dollar of cost for the taxpayers.

Transfers in the form of particular services rather than money nevertheless may be justified in extreme cases. For example, providing cash to drug-addicted parents is unlikely to assist their children. It should generally be presumed, however, that recipients know best how to use a given amount of money to make themselves better off. Transfers that restrict freedom of choice for the recipients therefore ought to be the exception to the rule. Furthermore, in order to limit the efficiency costs of in-kind transfers, there ought to be strong presumptions in favour of restricting subsidies on the basis of need and subsidising provision from competitive private suppliers, and against monopoly government supply.

DISCRIMINATION AND EQUITY

As we noted above, equity can also be taken to mean fair or equal treatment. People ought not to be victimised or to suffer from unreasonable discrimination. It is reasonable to discriminate against job applicants because they lack the skills needed for particular jobs. It is unreasonable to discriminate against job applicants because they lack or possess traits that are unrelated to their abilities to perform the jobs they seek.

Advocates of affirmative action argue that the notion of equity as equal treatment ought to be set aside when an identifiable group of people has suffered from past unreasonable discrimination that has adversely affected the welfare of its members. Some argue that positive discrimination can be used to offset the effects, or at least to compensate for the cost, of past negative discriminatory acts.

We argue that one-off income or resource transfers to members of a disadvantaged group are preferable to legalised discrimination. Specifically, if members in a particular group have been adversely affected by past discrimination, they will be eligible to obtain a disproportionate share of income transfers that are available on a needs basis.

Members of an identifiable subgroup of the population may also have disadvantages as a result of past discrimination that may not be easily corrected by income transfers. Again, there is no logical reason for barring people who need in-kind assistance from receiving it simply because they do not fall into some identifiable subgroup. By the same token, such assistance ought not to be provided to people within the subgroup who do not need it.

RECOGNISING COMPASSION, REJECTING ENVY

Promoting equality is a misguided, and potentially dangerous, goal. Policies aimed at promoting equality are likely to impose substantial costs in terms of lost prosperity and freedom and may also compromise other notions of equity, such as fair and equal treatment. In particular, policies aimed at appeasing feelings of envy ought to be rejected.

Limited government action is necessary, however, to assist people in need who fail to receive adequate support from family, friends or private charity. Such assistance is best provided through explicit subsidies of services purchased in a competitive market rather than through establishing monopoly government agencies. Assistance to those in need should be in the form of cash transfers unless a strong case can be made for alternative forms of help.

The distributional benefits of policies cannot be gauged adequately by measures that ignore the distinction between upside and downside inequality. The standard measures of inequality are fundamentally flawed. Such measures perpetuate policies that are misguided when viewed in terms of their stated goals, and possibly corrupt when their true purposes are revealed.

EQUITY AS FAIR ALLOCATION

Our discussion focuses on the notion of equity as a social goal. We are interested in questions such as the mobilisation and allocation of resources within society, the rights and obligations of individuals, and the nature of the governmental, legal and economic institutions that society ought to have.

We have little to say about an extensive literature on another notion of equity that arises when a decision-maker has to allocate some particular resources or burdens to a group of identified individuals. For example, a person writing a will may be deciding how to divide assets among the beneficiaries. A kidney available for transplant needs to be allocated to just one of the group of people who have been identified as suitable recipients. The fraction of time two parents can spend with their children needs to be determined as part of a divorce settlement.

There are a number of distinctions between this notion of equity, which Young (1994) also calls “local or micro justice”, and the one we shall be discussing, which he designates as “macro justice”. Young comments that the issues that are the focus of macro justice “do not tell us how to solve concrete, everyday distributive problems” and explains:

> The reason for this failure is that issues of local justice tend to be compartmentalised. Society makes no effort to coordinate distributive decisions across different domains, nor would it make sense to try … There is no mechanism comparable to the invisible hand of the market for coordinating distributive decisions at the micro level into just outcomes at the macro level. To the extent that gross inequities exist because of the cumulative effects of local decisions, they can be rectified at the societal level by redistributive policies, not by trying to coordinate the decisions of local allocative institutions. (Young, 1994: 6–7)

There are other differences between the micro questions on which Young focuses and the macro problems that concern us. In the type of questions we examine, there are no identified resources or burdens that are being allocated. One of the critical issues we consider is how different rules or institutions affect the overall availability of resources.

Furthermore, in the situations in which we are interested, decision-makers generally will not know how particular individuals will be affected by their decisions. Particular outcomes depend not only on general rules or institutions but also on many facts that are unknown, and are unlikely to become known, to the decision-maker. Moreover, the particular individuals who will be affected by a rule or institution are not known when the rules or institutions are chosen or devised. In this situation, the salient issue is the likely effect of a rule or institution on average and over the long run, not whether the outcome in one particular instance is considered fair or equitable.

In addition, and unlike the situations of fair allocation mentioned above, there may be no identifiable decision-maker who can be held responsible for any particular outcome. A modern economy is not a tribal society in which an identified decision-maker knows all the people affected by decisions relating to the allocation of resources or burdens.

The fair allocation mechanisms that Young (1994) discusses nevertheless are related to some of the principles we propose for justice at the macro level. In particular, Young argues for the relevance of principles such as efficiency, equal treatment of equals and stability and predictability in allocative rules. These are all principles for which we also argue at the level of rules governing social institutions as a whole.

2

ETHICAL FOUNDATIONS

The institutions and rules of social interaction that characterise a good society, and whether concerns about equity ought to be reflected therein, are the focus of this work. In this chapter, we consider a framework for evaluating social, political and economic institutions and rules. We also discuss the different notions of equity that will be examined in subsequent chapters.

WHAT IS EQUITY?

In its common use, equity refers to what is fair and just, as opposed to that which is merely legal. The term also encompasses the notion of natural justice and of application of a set of rules equally to all. When we speak of an equitable agreement, we mean that each side is treated fairly.

In ordinary usage, 'equity' and 'equality' are used interchangeably. Some claim that a society can be more equitable only if people are more equal in some sense. Policies that result in individuals being treated differently are said to be justified when they reduce inequalities.

An immediate difficulty is that there is no unambiguous measure to determine how well-off people are. The most comprehensive and defensible measure of a person's well-being would be lifetime expected welfare, or 'utility'. Unfortunately, as we discuss further in Chapter 5, a practical problem with the concept is that it is not even in principle measurable.

The opportunities that any person has to use time gainfully are in principle measurable, as is that person's access to market goods and services. Whereas utility is a subjective concept, the opportunities that a person can exploit in order to increase utility are objectively verifiable by an outside observer. Thus, as noted in Chapter 1, another concept of equality often discussed is equality of opportunity. Equality of opportunity is considered in more detail in Chapter 7, and we show that it also suffers from intractable problems related to measurement. In light of the difficulties in measuring inequality of welfare or opportunities, people concerned about inequality usually retreat to measuring inequality of annual money incomes or annual spending.

The fact that income inequality is not an ideal measure of inequality has important consequences. We should be cautious in implementing redistribution policies, since redistribution based on an imperfect measure might be counterproductive in equalising members of society. In their personal charitable giving, people support others who are experiencing material deprivation, but are less willing to help people who have access to basic material necessities. We suggest that the same caution is warranted in implementing public redistribution programmes.

Equality before the law differs markedly from equality of material wealth or equality of opportunity. Equality before the law concerns processes rather than outcomes. An equitable arrangement of social institutions is one in which individuals have the same right to pursue their own self-interest, provided they do not infringe on others' rights to pursue their goals. Laws apply equally to all members of society and no person is above the law. In contrast, identifying equity with equality in the distribution of income suggests that we are justified in imposing sacrifices on some in order to advance the welfare of others.

Promoting equality of income ought not to be a function of government, even though compassion for those in need is a laudable goal. Caring for one's friends and neighbours, helping them out in times of need, and making sacrifices on their behalf, are all worthy of praise. There may also be reasonable arguments for limited use of the coercive powers of government in the name of compassion. Using government powers compassionately is different from using them in an attempt to achieve equality. As Jouvenal noted:

> Let me only underline that to whatever extent we feel that the uplifting of the least favoured members of our society is called for, this is not logically bound up with the demand for equalization of incomes. The latter concept has been seen to lack any secure basis: It is unclear in idea and in its destructive aspect a transient rather than a fundamental feeling. The method of so-called redistribution through the agency of the redistributing State and its outcome, the favouring of corporate bodies over individuals, seem to us to pertain to a vast evolutionary process which will not result in equality, and in which the egalitarian ideal is put to work, in all good faith, for ends other than itself. (Jouvenal, 1952: 80)

While natural justice, or fair play, generally requires the application of a set of rules equally to all, in some circumstances it may be fair or just to make special allowances. People in dire circumstances could be assisted through income transfers, lower taxes, subsidies for the consumption of particular goods or services, or legal rules requiring people to be treated differently in market transactions.

Different treatment for individuals could be implemented with minimal infringement of fundamental rights by a basic state safety net, or means-tested vouchers for education, health care or other services considered fundamental to an acceptable standard of living. The existing role of the state in redistributing through the tax system and providing education, health and welfare services goes beyond that justified on the grounds of compassion.

A BRIEF OVERVIEW OF ETHICS

Our study focuses on what we ought to do as a political or communal entity, not on the ethical behaviour of each of us as individuals. Some problems that arise in our discussion of political philosophy, however, have deeper roots. They are, in fact, inherent in any discussion of ethics.

Problems inherent in any discussion of ethics

Ethics is concerned with what people ought to do. Ethics is a difficult subject because it straddles the theoretical and empirical realms. On the one hand, ethical principles are statements about ideal behaviour and are not based on the way people actually behave, since people do not always do the right thing.

It was wrong for Hitler to murder millions of people, regardless of his personal moral code or the fact that many people in Germany supported his actions or any other mitigating factors to which he might have appealed in defending his actions. In extreme cases such as this, few people would dispute the necessity and morality of action to defend a moral law. The truth of the moral law, in this case the law against mass murder, does not depend upon any particular empirical facts.

There is an important sense, however, in which ethical standards are empirical. When we assert that something is a moral law we are saying that following that law in a particular situation will lead us to take actions which we would intuitively judge to be the right ones under the circumstances. We test the law by examining a wide range of different situations and asking whether the action the law would specify in each case is also the action we would judge to be the right one. If not, we see if a new law can be found that predicts those actions we would regard as being right in a wider range of possible situations. The test is not whether the rule can describe actual behaviour in particular contexts. Rather, the empirical test is whether the rule would prescribe behaviour that one would regard as ethical in that context.

Similarly, one way to settle disagreements about competing moral laws is to examine what the competing laws imply for the right behaviour in different situations. By examining these cases we might hope to agree that one law is better than another. This will be a vain hope, however, unless we can agree on what is, or is not, appropriate behaviour in different contexts.

Some assert that there is no way of evaluating ethical rules. Rules are made up by each person or each society, and there is no way of reaching agreement on what might be better or worse ethical principles. If one ethical belief is as valid as another, then there is no point in arguing for or against any given ethical position. The outcome of any discussion is known in advance.

Similarly, some assert that ethical beliefs are axiomatic. For example, achieving equality is attested to be good in and of itself, and no discussion or argument could change the mind of someone who believes in this axiom.

Our work is based on the premise that we can reach agreement on desirable ethical rules by discussing the likely implications of different rules in particular contexts. We attempt to demonstrate in what follows that ethics is a useful subject of discussion. There are some ethical rules that are better than others, and our task is to discover them through a process of open dialogue. We may not know what the best principles are, and we may argue in favour of one set of principles and against another. If we want to discuss ethics rather than anthropology, however, we must begin with the premise that people of all cultural and historical backgrounds share some intuitions about right and wrong. They can agree about types of behaviour that should be considered ethical, and this shared intuition is the data we use to sort through competing ethical rules.

There is one ethical principle that ought to elicit universal assent. Let us call this the common sense principle: *If a person ought to do X, that person can do X*.

It would be nonsense to say that a person whose arms had been amputated ought to dive into a river to save a drowning person. ‘Ought’ implies ‘can’.

The common sense principle can modify our judgments about the morality of particular acts. It is often difficult to determine whether a person who commits a bad action could have refrained from committing it.

We cannot object when an amputee fails to save a drowning person, if the amputee also would have drowned after jumping into the river. Other cases are more difficult. If an insane person *A* kills another person *B* while *A* is hallucinating, we say that *A* could not have acted differently and that it does not make sense to say that *A* ought to have refrained from killing *B*. Suppose, however, that a person *C* who suffered from a bad upbringing, poor education and mild mental illness kills another person *D*. Some may deem that *C* ought to have refrained from killing *D*, while others deem that *C* could not have acted differently and, therefore, it does not make sense to say that

C ought to have refrained from killing *D*. Thus, a perennial problem for ethics is deciding the degree to which people have control over their actions.

The common sense principle has one important implication that is significant for political philosophy. The principle implies that if something cannot be done it makes no sense to say that it ought to be done. Political philosophers often specify what the goals of a society, or political entity, ought to be without discussing what is achievable in practice. For example, government responsibilities might be expanded to include policies aimed at achieving equality. In reality, however, the incentives inherent in democratic political systems may mean that the expanded powers would never be used for the intended purpose. If a proposed goal were unobtainable in any realistic setting, it makes no sense to maintain that governments ought to pursue it. The focus should shift instead to goals that are achievable.

Violations of the common sense principle are determined by empirical means. For example, whether one person or a group of people can act altruistically over a long period of time is an empirical question. The extent to which individuals can forgo immediate pleasures for the sake of long-term benefits is also an empirical question. More complex questions arise when we consider the interactions among people in society, and the way social institutions affect these interactions.

A fundamental problem in ethics is that there are often conflicts among the ethical principles that govern individuals and social institutions. The problem is determining which principles take precedence when a conflict arises.

On the one hand, we may value liberty, or individual freedom up to the point at which it infringes the rights of others. On the other hand, we may object to seeing a poor person lying in the gutter, starving to death, while a rich person has an ample store of food nearby. As Hume noted:

> Let us suppose a person ever so selfish, let private interest have engrossed ever so much his attention, yet in instances where that is not concerned he must unavoidably feel some propensity to the good of mankind and make it an object of choice, if everything were equal. Would any man who is walking along tread as willingly on another's gouty toes, whom he has no quarrel with, as on the hard flint pavement? … And if the principles of humanity are capable, in many instances, of influencing our actions, they must, at all times have some authority over our sentiments and give us a general approbation of what is useful to society and blame of what is dangerous or pernicious. (Hume, 1957: 52–53)

Although a respect for individual autonomy and property rights may deny us the right to take the rich person's food and give it to the starving person, some other ethical principle urges us to assist the poor.

There appears to be an inherent conflict between the goal of respecting individual liberty and the goal of acting compassionately. A respect for liberty implies that rights to one's person and property are secure and that individuals can allocate resources under their control as they wish. Compassion requires that we help those who are in desperate straits. If the wealthier segments of the community will not voluntarily help the less fortunate, then we may feel justified in compelling them to help.

When ethical values come into conflict, reasonable people can differ over the relative weights that should be given to each goal. Some of this disagreement may be a dispute over the relative merits of the different goals. Much of it, however, is likely to involve differences in judgment about the likely consequences of giving primacy to one goal or the other. Again, we are led to consider the relevance of empirical observations to the ethical positions we take.

THE SOCIAL CONTRACT

We noted that it is hard for individuals to arrive at reasonable personal ethical positions about the appropriate trade-offs to make among competing goals. A related problem arises in a social context. When reasonable people disagree about social goals, whose values should prevail? What are the social mechanisms for aggregating individual ethical judgments? Is there any guarantee that such mechanisms will give us desirable outcomes?

One assumption traditionally made is that each person (excluding children and those with mental illnesses) has an equal ability to determine what is just. For this reason, juries are composed of randomly selected citizens, each of whom is presumed equally fit to evaluate a case.

While it may not be true that people are equally well-endowed in telling right from wrong, to make the opposite assumption would cause intractable difficulties. There would be no way to decide who was the person endowed with the talent for identifying those with moral ability. Who could claim to possess, with any evidence, a superior moral ability to identify other morally superior people? We would be trapped in an infinite regress, for who would decree that the judge of the morally superior people was in fact a fit judge of such people?

As a practical matter, therefore, and in order to get people to agree to be constrained by a set of social rules, there is a strong presumption in favour of some type of democratic arrangement. Each citizen is presumed equally capable of judging appropriate social goals. Governments ought to govern only with the consent of the citizens, with each citizen having an equal say.

Even if we conclude that people ought to be given an equal say in determining social goals, we still need to arrive at a consensus concerning the rules that will apply to our decision-making process as a group. How will the equal say of each person be aggregated to make a decision that binds us all? One way of determining the best rule is to reason about the implications of different decision-making rules and to see if we can agree on which rules are likely to produce the best outcomes.

One simple example is used with children. Suppose we have a piece of cake to be divided between two children. We think it is most fair to allow one child to cut the cake and the other child to pick the first piece. The child doing the cutting does not know which piece will be left behind, and thus has an incentive to ensure that the pieces are cut fairly.

Some philosophers argue similarly that we can arrive at reasonable rules of social decision-making by asking how individuals who have not yet joined a society would act. How can a society make itself attractive to potential new members?

No one would agree to join a society if everyone in the group did not agree to certain rules. These rules would include constraining people from taking others' property and from injuring other people. A person or family living outside of a group would spend substantial time and effort fending off others who may wish to inflict harm, or to steal food and belongings. When people band together into a society, they can provide for their common protection against outside enemies, for internal law enforcement and for other jointly provided services much more effectively than can each person alone. While there may not be an explicit contract between the parties to this deal, there may be a type of implicit contract.

Similarly, individuals might voluntarily agree to be coerced into contributing to the supply of certain types of jointly used facilities or services. When the time comes to fulfil commitments, each individual might prefer to renege. It may nevertheless be rational to agree to such an arrangement *ex ante*, if others also agree to be coerced to contribute *ex post*.

Locke describes this original agreement, called the social contract, when he writes:

> Man, being born, as has been proved, with a title to perfect freedom and uncontrolled enjoyment of all the rights and privileges of the law of nature equally with any other man or number of men in the world, has by nature a power not only to preserve his property—that is, his life, liberty, and estate—against the injuries and attempts of other men, but to judge of and punish the breaches of that law in others as he is persuaded the offense deserves, even with death itself in crimes where the heinousness of the fact in his opinion requires it.

> But because no political society can be, nor subsist, without having in itself the power to preserve the property and, in order thereunto, punish the offenses of all those of that society, there and there only is political society where every one of the members has quitted his natural power, resigned it up into the hands of the community in all cases that exclude him not from appealing for protection to the law established by it. And thus all private judgment of every particular member being excluded, the community comes to be umpire by settled standing rules, indifferent and the same to all parties, and by men having authority from the community for the execution of those rules decides all the differences that may happen between any members of that society concerning any matter of right, and punishes those offenses which any member has committed against the society with such penalties as the law has established; whereby it is easy to discern who are, and who are not, in political society together. (Locke, 1952: 48–49)

Similarly, Nozick (1974: 3) begins his work on political philosophy by defending social contract theory as a way of reasoning about the goals and purposes of political activity:

> The fundamental question of political philosophy, one that precedes questions about how the state should be organised, is whether there should be any state at all. Why not have anarchy? …
>
> Which anarchic situation should we investigate to answer the question of why not anarchy? Perhaps the one that would exist if the actual political situation didn't, while no other possible political one did …
>
> Were this description awful enough, the state would come out as a preferred alternative, viewed as affectionately as a trip to the dentist. Such awful descriptions rarely convince, and not merely because they fail to cheer …
>
> More to the point, especially for deciding what goals one should try to achieve, would be to focus upon a non-state situation in which people generally satisfy moral constraints and generally act as they ought. Such an assumption is not wildly optimistic; it does not assume that all people act exactly as they should. Yet this state-of-nature situation is the best anarchic situation one reasonably could hope for. Hence investigating its nature and defects is of crucial importance to deciding whether there should be a state rather than anarchy. If one could show that the state would be superior even to this most favoured situation of anarchy, the best that realistically can be hoped for, or would arise by a process involving no morally impermissible steps, or would be an improvement if it arose, this would provide a rationale for the state's existence; it would justify the state.

The idea of the social contract encourages us to abstract from our own particular circumstances. When we contemplate the social contract, we consider those institutions and rules of social interaction that might be beneficial on average and over the long term. Maximising our personal, short-term interests at the expense of long-term social benefits does not produce good social policy and is the antithesis of ethical behaviour. Social contractarians therefore urge us to characterise a 'good' society as the type of society we would have first organised as we moved from a state of nature to a state of civilisation.

Locke and other advocates of the social contract approach admit that such agreements may never have occurred in practice. Civil societies evolved, rather than springing to life at a set moment in time. The idea of the social contract is, however, a useful thought experiment.

The social contract as a legal fiction

Social contractarians view all social institutions, including government, as modes of interaction that have been invented by people to achieve ends that cannot be achieved more effectively through other means. The goals, purposes and legitimacy of government actions are seen as being derived from the goals, purposes and moral precepts of the people who consent to be governed.

The social contract is an example of a legal fiction. A legal fiction is some aspect of law or government that we know is false, but we pretend is true. For example, it is impossible for anyone to know all the laws of a particular country. Yet such knowledge is presupposed when we claim that "ignorance of the law is no excuse". We accept as a legal fiction that everyone knows all the laws of the land, because to assume the opposite would lead to chaos. How could a court of law ever determine what a person knew at the time an alleged offence was committed?

As Holcombe notes:

> In ancient societies, the fictions of common ancestry and natural law were employed as a device to provide order in society and lend legitimacy to the actions of decision makers. The fiction of natural law also placed bounds on the permissible actions of political leaders. Common law in the English tradition employs the fiction that the law remains unchanged, with rulings in particular cases to be determined by the already existing laws embodied in past decisions, and democratic nations like the United States rely on a fiction that the majority rules in order to give legitimacy to state action. In fact, a small group of individuals determines American political policies and the vast majority have no say at all. (Holcombe, 1994)

The value of inherited rules and institutions

The legal fiction of a social contract provides a bridge between the normative and positive aspects of democratic government. On the positive or empirical side, one might argue that people living in democratic countries do, in fact, appeal to the authority of ancient contracts (such as the Magna Carta or the United States Constitution) and ancient writers who interpreted such contracts (such as Blackstone who wrote his *Commentaries* on English law or the authors of the Federalist Papers, a work that is used to interpret the United States Constitution).

More generally, many commentators have argued that inherited social institutions and rules may have value simply because they have survived, and been modified over time, in response to competition from alternative rules or institutions. We may not even understand fully why certain inherited rules or institutions have been successful. For example, Hayek observes:

> Like all general purpose tools, rules serve because they have become adapted to the solution of recurring problem situations and thereby help to make the members of the society in which they prevail more effective in the pursuit of their aims. Like a knife or a hammer they have been shaped not with a particular purpose in view but because in this form rather than in some other form they have proved serviceable in a great variety of situations. They have not been constructed to meet foreseen particular needs but have been selected in a process of evolution. (Hayek, 1976: 21)

Rules of conduct that are detrimental to economic prosperity and social harmony may be less likely to survive because the societies they produce are less likely to survive or to be imitated. In particular, the appeal to precedent yields stability and predictability in the political and legal systems, which in turn is of fundamental importance to investment and economic prosperity.

An implication of Hayek's perspective is that a degree of humility may be in order when proposing new policies or institutions. On this point, Hayek is in good company. For example, Smith noted:

> The man whose public spirit is prompted altogether by humanity and benevolence, will respect the established powers and privileges even of individuals, and still more those of the great orders and societies, into which the state is divided. Though he should consider some of them as in some measure abusive, he will content himself with moderating what he often cannot annihilate without great violence. When he cannot conquer the rooted prejudices of the people by reason and persuasion, he will not attempt to subdue them by force … When he cannot establish the right, he will not disdain to ameliorate the wrong … when he cannot establish the best system of laws, he will endeavour to establish the best that the people can bear.
>
> The man of system, on the contrary, is apt to be very wise in his own conceit; and is often so enamoured with the supposed beauty of his own ideal plan of

> government, that he cannot suffer the smallest deviation from any part of it. He goes on to establish it completely and in all its parts, without any regard either to the great interests, or to the strong prejudices which may oppose it. He seems to imagine that he can arrange the different members of a great society with as much ease as the hand arranges the different pieces upon a chessboard. He does not consider that the pieces upon the chessboard have no other principle of motion besides that which the hand impresses upon them; but that, in the great chessboard of human society, every single piece has a principle of motion of its own, altogether different from that which the legislature might choose to impress upon it. (Smith, 1963: 410–411)

Because of the difficulty in knowing whether and how people can adapt to a new situation, Smith argues that we should undertake social change carefully. Making small changes to an existing, imperfect social institution may be preferable to attempting the creation of a new, more perfect one from scratch.

Respect for inherited rules and institutions gives people some certainty about the decisions that will be made by others both now and in the future. Individuals can then make their own decisions, and plan their own future actions, with greater confidence. People are more likely, for example, to make contractual agreements when they know how such contracts are likely to be interpreted by the courts. If people fear that radical shifts in government policy will significantly change future costs or benefits, they will be wary of saving and borrowing, investing, entering into insurance contracts and making myriad other forward-looking decisions. Such decisions are, however, crucial for the accumulation of capital and thus for increasing economic prosperity.

When decisions are not constrained by rules, whatever is expedient becomes legitimate. A famous example of such a disaster was the French Revolution. French intellectuals argued rightly that there was no logical reason why the habits and institutions of France's past should be the ones pursued by France in the future. Pure reason alone, they contended, could devise better governments, better social institutions, and better laws, as well as better calendars and systems of measurement. The results of the French attempt at revising the social order are well known. Thousands of innocent people were killed and the revolution set the stage for a war that ravaged a large portion of Europe. Writing before the full effects of the French Revolution were known, Burke noted:

> The science of constructing a commonwealth, or renovating it, or reforming it, is, like every other experimental science, not to be taught *a priori*. Nor is it a short experience that can instruct us in that practical science; because the real effects of

> moral causes are not always immediate; but that which in the first instance is prejudicial may be excellent in its remoter operation; and its excellence may arise even from the ill effects it produces in the beginning. The reverse also happens; and very plausible schemes, with very pleasing commencements, have often shameful and lamentable conclusions. In states there are often some obscure and almost latent causes, things which appear at first view of little moment, on which a very great part of its prosperity or adversity may most essentially depend. The science of government being therefore so practical in itself, and intended for such practical purposes, a matter which requires experience, and even more experience than any person can gain in his whole life, however sagacious and observing he may be, it is with infinite caution that any man ought to venture upon pulling down an edifice which has answered in any tolerable degree for ages the common purposes of society, or on building it up again, without having models and patterns of approved utility before his eyes. (Burke, 1986: 152)

Preserving the status quo has value because the costs of change are high and the results of change are unpredictable. This is not to imply, however, that laws and social institutions should never be changed or that, as Epstein (1999: 282) suggests "conscious acts of statesmanship by analogy, as it were, to breeding and other forms of artificial selection" cannot improve things. Indeed, with regard to the United States Constitution Epstein notes:

> The ideal constitution within a Hayekian world would be the English Constitution with its unclear origins, slow growth, and shadowy conventions. But he would be suspicious of the quite deliberate architecture of the United States Constitution, with its conscious reliance on the staples of political theory – separation of powers, checks and balances, enumerated powers, federalism, and entrenched rights …
>
> Indeed it is not far-fetched to see in the United States Constitution a rejection of the idea that a sound government could just evolve from the uncoordinated activities of well-intentioned individuals, or even thirteen somewhat fractious American states. (Epstein, 1999: 282)

Furthermore, when we move beyond discussing constitutional rules to actual policies we must be prepared to learn from past mistakes, and to be open to the possibility that new information or knowledge could change our assessment of the likely effects of different policies. In addition, rapidly evolving technologies and other social or environmental changes may also require us to alter many of our institutions and policies.

New Zealand has benefited from the reform process undertaken in the last decade-and-a-half relative to how it would have fared if such reforms had not been implemented. A standard principle of economics, supported by evidence from many countries and eras, is that tariffs and other restrictions on trade, or subsidies for particular industries, are more likely to reduce than to increase the average standard of living for residents of a small open economy such as New Zealand. Similarly, the two decades following the mid-1960s provided evidence that activist fiscal and monetary policy does more harm than good. Finally, the results of the wave of privatisations and deregulation that swept the world in the 1980s demonstrated that government-owned business enterprises perform poorly when measured in terms of minimising the costs of production or providing service to customers (see, for example, Bishop and Kay (1988), Boardman and Vining (1989), Shirley and Nellis (1991), Kikeri, Nellis and Shirley (1992), Galal *et al* (1994) and Megginson *et al* (1994)).

While there is much to be said in favour of old rules, particularly those that have proved successful, new arrangements can improve matters. We are left with the problem of choosing reasonable ethical criteria by which we might evaluate the social utility of new rules or institutional arrangements. One way of finding such criteria is by examining the ethical theories that underlie our everyday moral decisions.

TWO POPULAR ETHICAL POSITIONS

There are two types of basic ethical theory which, in various guises, are endorsed by many modern scholars. The first is the deontological or Kantian-type ethical theory. The second is consequentialism. These theories start from different premises, yet they often lead to the same sorts of practical rules for human behaviour.

Our consideration of these views is by necessity brief and incomplete. The key distinction is between rules, or constraints on choices, and choosing between alternatives based on their consequences in a particular instance. This distinction may be particularly important in a social context because expectations about the likely behaviour of other people help determine the best action for any one person. While we argued above that there is a presumption in favour of a democratic form of government, there may also be arguments for constraining the types of actions that a democratically chosen government is allowed to undertake.

Kant's view

It would be convenient if we could derive our notions of justice and equity from first principles, without having to know particular facts about the world. Kant maintained that such *a priori* knowledge was both possible and essential. Kant held that the hallmark of rationality is consistency. As rational agents, we should adopt rules governing our own behaviour that we would like to see binding on everyone's behaviour. This is the notion of the categorical imperative for which Kant is famous. Kant maintained that people should ask themselves before performing any action, whether what they intend to do should be universalised, enshrined in a law by which all rational agents who find themselves in similar circumstances would be bound. (Kant, 1964)

A second important tenet of Kant's position is that rational agents should always be viewed as ends-in-themselves and never as means to other ends. We should never trample on one person's rights in order to advance another person's ends.

Kant attempted to establish a moral principle (the categorical imperative) that is true in the same way as principles of Newtonian physics were thought to be true at the time Kant was writing. In particular, Kepler's laws governing the motion of planets around the sun in our solar system were regarded as truths about the world (synthetic), but they were derived from mathematical truths (calculus) that would be valid whether any physical world existed or not. These are the so-called 'synthetic *a priori*' truths of science.

Moral truths, for Kant, are *a priori* truths, that is, they are true in all times and places, and they apply to all rational creatures, including, in principle, rational aliens from outer space. Kant defends his task of searching for *a priori* truths in morals when he writes:

> Since my aim here is directed strictly to moral philosophy, I limit my proposed question to this point only: Do we not think it a matter of utmost necessity to work out for once a pure moral philosophy completely cleansed of everything that can only be empirical and appropriate to anthropology? That there must be such a philosophy is already obvious from the common idea of duty and from the laws of morality. Every one must admit that a law has to carry with it absolute necessity if it is to be valid morally – valid, that is, as a ground of obligation; that the command 'Thou shalt not lie' could not hold merely for men, other rational beings having no obligation to abide by it – and similarly with all other genuine moral laws; that here consequently the ground of obligation must be looked for, not in the nature of man nor in the circumstances of the world in which he is placed but solely *a priori* in the concepts of pure reason; and that every other precept based on principles of mere existence – and even a precept that may in a certain sense be considered universal, so far as it rests in its slightest part, perhaps only in its motive, on empirical grounds – can indeed be called a practical rule, but never a moral law. (Kant, 1964: 57)

For Kant, human beings are by nature rational creatures. To be rational, for Kant, is to obey the laws of logical thinking, and the hallmark of logical thinking is consistency. Thus, ethical human behaviour must be consistent.

Kant interprets the requirement of consistency to mean that moral laws should be applied equally. Thus, if a course of action is moral for me, it must be moral for anyone in a situation similar to mine. All people must will that their actions become universal moral laws for rational beings, if such actions are to be moral.

A person's motivation for action, according to Kant, must not come from feelings or emotions, but solely from reverence or respect for the moral law. Thus, one should choose moral actions on purely rational grounds in order to fulfil one's natural role as a rational creature.

There are many problems with Kant's ethical position. One problem is that human beings seem to need a spark of emotion or feeling in order to be motivated to act. We are animals seeking gratification of desires, and it is unrealistic to expect ordinary people to govern all their actions by the light of pure reason.

A second important problem with Kant's position is that it is extremely rigid. His position does not accommodate situations in which one person's rights might justifiably be trampled upon for the sake of the greater good. Two examples may help clarify this point.

Suppose there were a woman who had a rare gene that rendered her immune to cancer. Call her Ms *X*. If *X*'s gene were transferred into other members of society (through genetic engineering techniques), the people receiving the transfer would also be immune from cancer. Suppose further that all *X* has to do in order to benefit mankind is to give away one strand of hair or one drop of saliva. Suppose we were willing to pay *X* a large sum of money for her hair or saliva. Suppose that, in spite of our generous offer, *X* refused to give us a sample of saliva or hair.

On Kant's view, we could not forcibly take the hair or saliva from *X*, because then we would be using *X* as a means to an end and not treating *X* as an end in itself. Even though the benefits to mankind would be enormous and the cost to *X* minuscule, we would not be morally justified in obtaining *X*'s genetic material without her permission. This seems to be an absurd result.

A second example concerns lying. Kant urges us to tell the truth, regardless of the consequences, because honesty is more important than any other ethical principle. On Kant's view, lying is, by its very nature, engaging in behaviour that is inconsistent, and consistency is essential in a rational creature. Of course, there are many cases in which we would consider lying to be the morally correct course of action. As Plato noted in *The Republic*:

> Suppose that a friend when in his right mind has deposited arms with me and asks for them when he is not in his right mind, ought I to give them back to him? No one would say that I ought or that I should be right in doing so, any more than they would say that I ought always to speak the truth to one who is in this condition. (Plato, 1973: 13)

Although Kant's position does have these and other problems, it has one attractive feature that is lacking in some other ethical theories. For Kant, there are absolutes in moral judgment. Some activities are right and others are wrong. The rightness or wrongness of the activities does not depend upon their outcome, but solely upon the intention with which they were performed. Kant's theory is an example of what is called a non-consequentialist ethical theory; evaluation of the ethical status of an action does not depend upon the outcome of the action.

There are clearly some aspects of non-consequentialism that most people endorse. For example, suppose that a man deliberately attempted to poison his aging mother in order to collect on her life insurance policy. Suppose further that the poison just happened to be the substance that restored her to perfect health. If the son's action were exposed, he could be charged with attempted murder, regardless of the beneficial effects of the poison on his mother. The fact that the son intended to kill his mother and took deliberate and rational steps in the direction of killing her makes him guilty of a crime. As we shall see, a utilitarian might view the attempted poisoning case somewhat differently.

Utilitarianism

There are as many consequentialist ethical theories as there are deontological or Kantian-type theories. We shall attempt here to describe in broad outline the subset of consequentialist theories known as utilitarianism.

A central tenet of utilitarianism is that more happiness (or utility) is better than less happiness. People generally seek pleasure and avoid pain, although they can endure short-run pains and deprivations in order to secure long-run benefits. Stating that people plan and carry out their lives in ways that they believe will maximise their own happiness is not equivalent to saying they are selfish. There is no necessary connection between utility and self-regarding behaviour. Utility could be increased, for example, by helping others in need.

Many versions of utilitarianism presuppose that the happiness of individuals can be aggregated into an unambiguous measure of social happiness or welfare. Utilitarians then claim that a society will be more ethical if it is organised so as to promote more rather than less social happiness.

A potential problem with utilitarianism immediately springs to mind. We might be able to make lots of people very happy if we are willing to make one person quite unhappy. For example, suppose that there were a homeless man Mr *Y*, living under a motorway who had no living relatives or friends. Suppose further that he is mentally defective but otherwise in reasonably good health.

Now suppose that in a hospital near *Y*'s motorway home there are many people awaiting organ transplants. Some of these individuals are medically suitable recipients for *Y*'s organs. Suppose that one is a world-famous painter who needs a heart. A second is a female research scientist who is on the verge of curing AIDS, but needs a kidney herself. A third is the Pope, a person who provides spiritual solace to millions, but needs a new liver.

We could take *Y*'s vital organs and transfer them to these deserving individuals on the organ transplant waiting list. The happiness of the world would be greatly increased, while the loss of *Y* would be relatively small by comparison. If we endorsed unrestrained utilitarianism (maximising aggregate 'social benefits'), we would be morally justified in killing *Y*. This is surely an unpalatable result.

Most versions of utilitarianism, including that first formulated by Mill, ascribe to individuals certain basic or natural rights. We may not violate these rights, even if violating them would lead to greater happiness for society as a whole. Then, subject to the constraint that we do not violate these basic rights, we should try to maximise the happiness of society as a whole.

We argue in Chapter 5 that utility is impossible to quantify, and hence that the concept of social welfare, or happiness of society as a whole, is of dubious merit. We can nevertheless analyse institutions and policies from a utilitarian perspective by assuming that most humans are rational, self-interested individuals. If people are given greater freedom to choose, they ought to be able to make themselves better off. They can always remain equally well off by not taking advantage of the new opportunities that have been made available to them. There is thus a presumption that greater individual freedom of choice and action is likely to be valued favourably by a utilitarian. Difficult cases arise where greater freedom of choice for one individual comes at the expense of less freedom of choice for another. Decisions of individuals may also lead to more desirable outcomes if they are coordinated. The choices available to any one individual, and the desirability of different actions one can choose, may depend on expectations about the behaviour of others. Legal, political and social institutions and rules may be of value because they help resolve some of these conflicts and coordination problems.

Pareto optimality and efficiency

Utilitarianism provides the normative foundations for economic discussions of public policies. Specifically, the implicit normative assumption underlying most economic analyses of public policies is that efficiency is desirable.

Efficiency measures the extent to which resources are being used to maximise community happiness, ignoring the distribution of benefits among people. More specifically, an allocation of resources is said to be *Pareto optimal* if the resources cannot be transferred to other uses to make someone better off without making anyone else worse off.

A simple example (inspired by trades apparently made in Eastern Australia for thousands of years) might clarify the concept of Pareto optimality. Suppose there are two people who both value rocks that can be used to make stone tools and ochre that can be used to decorate one's body. Person *A* lives in an area that has many fine sources of rock for stone tools but no good source of ochre. For person *A*, rocks are not very valuable but good ochre is of great value. Person *B*, on the other hand, lives in a desert area where clays can be found in many shades of red, orange, yellow, white and so on but where rocks are scarce. For person *B*, rocks are much more valuable than clay. Allowing *A* and *B* to trade with each other allows them both to become better off. Person *A* gives up relatively cheap rocks and obtains in exchange valuable ochre. Person *B* gives up relatively cheap ochre and obtains in exchange valuable rocks. If it were costless for these people to trade, they would continue trading until they had the same relative valuations of rock and ochre. The final allocation of rocks and ochre would be Pareto optimal. Neither person could be made better off without making the other worse off. Suppose, however, that meetings between *A* and *B* occur haphazardly. Person *A* then informs *B* of a group of bunya pines in A's area that produce edible nuts, and they arrange to meet each year when the nuts are ripe. If both *A* and *B* agree to the new arrangement, presumably it makes them both better off, and it would therefore be a pure Pareto improvement.

In more complicated examples, such as government policy changes in a country like New Zealand, it will usually be impossible to avoid making someone worse off. We rely instead on a less restrictive notion than Pareto optimality.

An allocation is said to be *efficient* if resources cannot be reallocated in a way that would allow the gainers to be better off after compensating any losers to ensure that the latter are no worse off. Under reallocations that increase efficiency, the gainers do not necessarily have to compensate the losers.

The distinction between Pareto improving, Pareto optimal and efficient resource allocations may be clarified with the aid of Figure 2.1 (modified

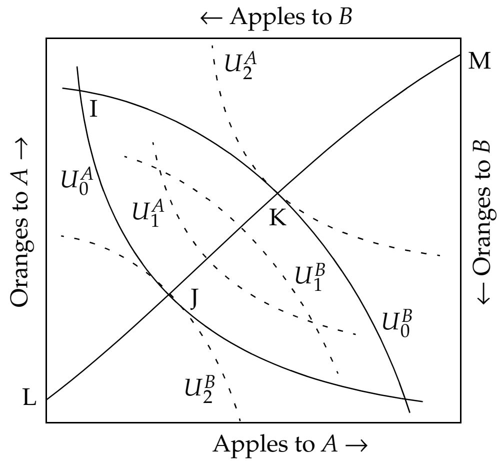

FIGURE 2.1: Pareto optimality and efficiency

from Coleman, 1980: 514). This figure represents the valuations by two people A and B of different allocations of a fixed total quantity of two goods, for example apples and oranges. The width of the box represents the total quantity of apples available to be divided between A and B, while the height of the box represents the total quantity of oranges to be divided between A and B. The *indifference curves* of person A (three representative ones labelled as U_0^A, U_1^A and U_2^A have been included in Figure 2.1) represent combinations of apples and oranges that leave A equally well off. They are drawn so that A's utility increases as allocations move away from the bottom left-hand corner of the box and A gets more of either or both types of fruit. The indifference curves of B (again three representative ones have been included in the figure) are drawn so B's utility increases as allocations move away from the top right-hand corner of the box, where A has all the available quantities of both types of fruit. Both sets of indifference curves are "bowed toward the origin" to reflect an assumption of decreasing marginal utility from increasing consumption of any one good. The more apples that *A* is already consuming, the greater the number of apples *A* would need to receive in exchange for one orange if *A* is to remain equally well off.

The initial allocation of fruit between *A* and *B* has been labelled as point I, at the intersection of indifference curves U_0^A and U_0^B. A reallocation of fruit from point I to any point inside the lens-shaped region bounded by curves U_0^A and U_0^B would represent a Pareto improving allocation. At least one of *A* or *B* must be made better off by such a reallocation and neither would be worse off. The Pareto optimal set of allocations relative to I is the curve JK of points of tangency between the two sets of indifference curves that represent at least as much utility as U_0^A and U_0^B. The points of tangency are significant because a lens-shaped region of possible Pareto improving gains will be absent only if the allocation is at such a tangency point.

The set of efficient points is, in this example, also known as the *contract curve* LJKM. This curve is the set of all points of tangency between the indifference curves of *A* and the indifference curves of *B*. Only if the current allocation is on this contract curve would it be impossible to find a reallocation that would enable the gainers to compensate the losers and still be better off. The set of efficient allocations includes LJ and KM where one of *A* or *B* is made worse off relative to the initial allocation I.

If those who gained always had to compensate those who lost, it may in practice be very difficult to implement policies that increase efficiency. A losing party would have an incentive to hold out on approving the change until all the surplus resulting from the improved allocation was paid as compensation. Furthermore, it would be very difficult, if not impossible, to identify all the losing parties from a policy change, and if compensation were being offered no doubt a large number of people would step forward to claim such a status.

The normative appeal of efficiency as a social goal is that an inefficient allocation of resources represents, in a sense, 'waste'. If someone could be made better off to a greater extent than others are made worse off, why not make the change? More explicitly, we argue in Chapter 3 that pursuit of efficiency is likely to be included, along with other goals, in a social contract formed by rational self-interested individuals.

Coleman (1980: 545–548) criticises the notion that the pursuit of Pareto optimality would be the sole feature of a social contract formed by rational self-interested individuals. The essence of his objection is that rational self-interested individuals would be concerned about the distribution of benefits resulting from economic activity. In the example presented in Figure 2.1, *A* and *B* would care about more than ensuring the economy ended up somewhere on the contract curve LJKM. They would also like to include some features in a social contract that limited where along LJKM the economy might end up. They might even be willing to countenance having the economy end up off the contract curve if they might thereby ensure more desirable distributional outcomes.

Many economists temper their support for efficiency by recognising that respect for individual liberty and some notion of equity may also be important determinants of the desirability of different policies. Even when other goals are being pursued, however, efficiency remains relevant as a measure of how much prosperity the society is forgoing on average in order to further other goals. Chapter 3 contains further discussion of the normative basis for including efficiency among the goals to be pursued by society.

Horizontal and vertical equity

One objection to the notion of efficiency is its distributional agnosticism. We might modify the extent to which we pursue various changes aimed at increasing efficiency, given our assessment of how those changes may affect the distribution of the benefits obtained from economic activity.

Economists state that a law, or set of legal or political institutions, is *horizontally equitable* if it treats like individuals in the same way. Discriminatory laws or policies that differentiate between people on the basis of unequal treatment based on irrelevant characteristics, are undesirable.

An example might help clarify the concept. Proposition 13, passed by referendum in California in 1978, limited *ad valorem* taxes on real property to 1% of the assessed property value in 1975–76. The adjusted values could then be increased by no more than 2% per year as long as the same taxpayer continued to own the property. For property sold or newly constructed after March 1, 1975, the assessed value was set at the appraised (or market) value at the time of sale or construction. A major objective of the initiative was to limit spending by local governments through restricting their access to revenue. Studies have suggested, however, that the law has increased the disparity in property tax liabilities for neighbouring properties of similar market value. There are cases in which one homeowner might pay more than 10 times the property taxes of a next door neighbour in a similar house. A recent study by two economists (Sheffrin and Sexton, 1998) found that a new homebuyer in Los Angeles County paid almost four times the average basic property tax paid by owners who had occupied their houses since 1975. Such a state of affairs strikes most people as unjust.

Economists have pointed out that the property tax regime in California results in a number of inefficiencies. Because people are likely to pay more property taxes if they change residence, the regime discourages moving. If the location of a job changes, people may prefer to stay where they are and commute a long distance rather than move. Many more people drive from homes to places of work than one would find if there were no tax on moving. Similarly, people are discouraged from moving out of their houses as family circumstances change. The result has been that many larger homes are occupied by smaller family groups, or more generally, many people are living in types of accommodation that they do not really prefer.

Inefficiencies may be part of the reason people object to the property tax differences in California. We suspect, however, that more is at stake. Such stark differences in tax payments resulting from the number of times a property has been sold, but unrelated to the value of the property, violate the fundamental rule that equals should be treated equally.

Consider also the following example. Suppose it were feasible to ascertain a person's genetic endowment at low cost. Assume further that the government of the day decided to charge lower income taxes for people who happened to be endowed with certain genes. If we rule out genetic engineering and tax-motivated mate selection, resources would not be wasted by people trying to adjust their genes (or those of their children) to minimise the tax burden, for such adjustments would be impossible. Resources may still be wasted, however, as different groups lobbied to have their particular genes treated favourably under the tax code. Nevertheless, this would probably not be the main objection to the policy. Rather, the policy would strike most people as offensive on the grounds of fairness. People in the same circumstances, in this case earning the same income, would be paying different amounts of income tax.

It is not a simple matter, however, to apply the concept of horizontal equity. In particular, when ought we to regard individuals as being 'alike' or, equivalently, when are differentiating characteristics 'irrelevant'? If we say simply that all people ought to be equal before the law, then horizontal equity corresponds closely with the notion of equity or fairness that is described by ancient philosophers such as Aristotle and Cicero and by more modern philosophers such as Hobbes, Locke, and Hume. When we examine issues such as tax policy, however, it is usually regarded as horizontally equitable to treat people differently but only in a limited sense. If the tax applies to a particular base, such as income, then people with the same income ought to pay the same tax. It would not be regarded as a violation of horizontal equity, however, for people with different incomes to pay a different amount of tax. In practice, the intended tax base may be difficult to determine. For example, should deductions for dependants be allowed in calculating the base for an income tax? Some of these issues are discussed further in Chapter 5.

It is also important to note that horizontal equity as we have defined it applies to laws or institutions. The concept focuses on processes, not particular acts or outcomes. A law or institution ought to treat people differently only if they differ along certain specified dimensions. If other characteristics are allowed to influence whether or not the law applies, or the extent to which it applies, the law can be made to apply to anything at all.

As noted earlier, there is also a notion of equity in modern usage that involves achieving equality of outcomes for different people. The characteristics of some people are said to justify different treatment, but only in so far as such different treatment reduces inequality of outcomes. This corresponds to the economist's notion of *vertical equity*, and is what we are referring to as equality.

Some commentators argue that we cannot achieve horizontal equity without a degree of vertical equity; that is, in a society in which the gap between the rich and the poor is very wide it will be impossible to enforce even the most ideal set of rules regarding fair play and equitable treatment of equals. Some also argue that a set of laws may be fair in principle, but often applied inequitably in practice, due to vast differences in wealth among certain classes of society. For example, in a society in which there are vast discrepancies in wealth, the wealthier individuals can buy better quality lawyers and secure better outcomes in a court of law. The wealthier can, however, also buy better quality education, health care, food, shelter and many other goods and services that are considered essential.

We argue later that the critical issue is ensuring that people do not fall below a minimal acceptable standard of living. In our view, inequalities among people, each of whom has a decent standard of living, are not important enough to justify the large efficiency costs of redistribution.

Efficiency and individual freedom

One could argue that freedom and horizontal equity are Kantian-type notions. We may affirm the importance of liberty and equal, or non-discriminatory, treatment as ends-in-themselves and not means to other ends, such as efficiency. A concern for liberty leads one to value the process of market exchange more than the results it produces. The emphasis is on freedom to exchange with whomever we wish, on ease of entry into markets, and on voluntary cooperation over coercion.

A concern for liberty also influences attitudes toward activities often considered beyond the scope of economic analysis. The coercive power of governments ought to be used as a means of last resort to achieve benefits that cannot be achieved voluntarily. People should be free to pursue their own moral or religious values or cultural practices provided they do not infringe on others' rights. By the same token, the coercive power of the state ought not to be used to promote particular moral, religious or cultural values.

From an economic perspective, promoting efficiency and respecting individual freedom and equal treatment may, in practice, often lead to similar policies or rules. This is more likely to be so if one insists on strict Pareto improvements rather than efficiency as we have defined it. Recall that under a strict Pareto improvement, no one can be made worse off. Such a constraint therefore requires us to respect individual rights.

Even under the definition that requires those who lose to be compensated, however, liberty and efficiency may conflict when people care about the actions of others. Interdependencies of utility are not a problem for efficiency if the concern of any person *A* for any other person *B* is based on *B*'s welfare as perceived by *B*. The conflict arises when the utility of one person depends directly on the choices made by someone else and the two people rank those choices differently.

For example, if parents care about the utility of their children, allowing the children to maximise utility will also make the parents better off. If parents believe that children do not know their own best interests, however, a conflict will result. The children would prefer to make one set of choices, while the parents may prefer that the same resources were used in a different way. A concern for liberty may lead us to respect the preferences of the children. A concern for an efficient allocation of resources would lead us to take the preferences of both the parents and the children into account.

Resource allocation within families is usually left to the discretion of family members. The presumption is that negotiation costs among family members are low, while close and continual contact allows people to learn the true preferences of other family members. Even if the outcome chosen by a family is not ideal, it may be the best among those that are feasible. Third parties will find it difficult to obtain the knowledge of preferences and opportunities that would be required to improve upon the choices made by the family.

The situation is more problematic, however, when interdependencies exist among strangers. Consider the case of a charitable donation when the donor has preferences for the way the money transfer is spent by the recipient. Efficiency would require us to take account of the wishes of the donor. A concern for individual freedom may urge us to allow the recipient to choose how to spend the money. A liberal donor would decide only whether a person deserves material support. The recipient would be left to decide how best to allocate that support to different needs or desires.

Another situation in which liberty and efficiency may conflict involves the use of government coercion to provide public goods, or to correct other so-called 'market failures':

> Where the conditions of competitive equilibrium are not satisfied, a market will not secure an efficient outcome. To secure the efficient outcome, the political order must intervene in the market. (Coleman, 1980: 542)

The conclusion Coleman draws does not follow from the premise. Although markets may not secure an efficient outcome, it does not follow that political mechanisms can do any better. A failure of markets to produce an efficient outcome provides the *potential* for alternative institutions to increase efficiency. Whether any actual political institution can achieve that potential gain is something that can only be determined by examining how political institutions operate. Thus, as we argue further in Chapter 3, the social contract formed by rational self-interested individuals would be informed by people's assessment of the likely consequences of using different institutions to perform different tasks.

A similar objection applies to another of Coleman's arguments that suggests a potential conflict exists between liberty and efficiency:

> … lack of adequate information may transform free choice into something less than Pareto-superior action. Further, it is at least plausible that some individuals acting freely make themselves worse off; freedom does not necessarily ensure increased happiness. (Coleman, 1980: 541)

Again, the issue is not only whether free choice leads to efficient outcomes but also whether there is any practical alternative that can do any better. In particular, there is no guarantee that better outcomes can be achieved by allowing government officials monitored by democratically elected politicians to make choices on behalf of individuals.

Nevertheless, we argue in later chapters that at least some collective action is likely to increase efficiency relative to unrestrained individualism. Indeed, a society based on markets and individual freedom would be impossible without a monopoly enforcer of the rule of law. Whenever coercive political action may be warranted on efficiency grounds, there can be a conflict between liberty and efficiency. Individuals may object to paying taxes, but it may be more efficient to force them to do so if that allows a more efficient level of public goods to be provided. In these cases, we might argue that efficiency ought to prevail over liberty.

Hayek's critique of utilitarianism

In light of the prominent role that efficiency plays in economic analyses of policy, it is surprising that one of the great defenders of market liberalism was critical of utilitarianism. For Hayek, the most powerful argument against socialist central planning was that no one person, or government organisation, could possibly know, let alone successfully utilise, the mass of detailed and local knowledge upon which millions of daily economic transactions rely. The same perspective led him, however, to reject a "narrow form of utilitarianism":

> The fact, however, is that in a Great Society in which the individuals are to be free to use their own knowledge for their own purposes, the general welfare at which a government ought to aim cannot consist of the sum of particular satisfactions of the several individuals for the simple reason that neither those nor all the circumstances determining them can be known to government or anybody else. (Hayek, 1976: 2)

One could view Hayek's argument as an application of the common sense principle. A government ought not to have as a goal something which it cannot achieve, and a government cannot achieve particular end states because it lacks the incredible amount of information that would be necessary to do so. Furthermore, there is no conceivable mechanism for a government ever to obtain access to the required information.

Hayek argues that utilitarianism is based upon the notion that we can create rules that will maximise social utility. In fact, however, when we create new rules, we do not know the effect they will have in particular cases. Our ignorance of the particular effects of laws, Hayek says, is the very reason we insist upon their being obeyed regardless of the circumstances.

The likely consequences of different rules must play some role, however, in determining their desirability. Indeed, Hayek eventually declares himself in favour of something similar to efficiency:

> We may of course aim at the 'greatest happiness of the greatest number' if we do not delude ourselves that we can determine the sum of this happiness by some calculation, or that there is a known aggregate of results at any one time. What the rules, and the order they serve, can do is no more than to increase the

> opportunities for unknown people. If we do the best we can to increase the opportunities for any unknown person picked at random, we will achieve the most we can, but certainly not because we have any idea of the sum of utility of pleasure which we have produced. (Hayek, 1976: 23)

The key distinction here is between predicting the specific outcomes of a set of rules and the general tendencies they imply. It is undoubtedly true that much of the information that would be required to predict the specific consequences of new rules, including the more remote and indirect effects, is widely dispersed. There is no known mechanism for truthfully revealing such information to a single decision-maker. Yet economists rarely make claims about the specific consequences of proposed policies. Rather, it is usually asserted that certain policies, institutions or social arrangements are more likely to achieve an efficient allocation of resources than are the alternatives.

For example, there is no way of knowing whether an allocation of resources is Pareto optimal simply by examining who owns what. There is instead a presumption that if two people voluntarily exchange one resource for another, each person expects to be better off as a result of the trade. If they are free to do so, people could be expected to trade with each other until the potential gains are fully exploited. Our knowledge of the way people will behave in response to incentives and constraints leads us to expect efficient outcomes when there is freedom to trade.

Our inability to predict the specific consequences of different policies is another reason economists focus on efficiency, as we have defined it on page 29, rather than strict Pareto optimality. There is simply no way of determining everyone who gains and loses from particular policy changes, nor the extent of their gains and losses. Rather, the claim that alternative policies, institutions or rules would increase efficiency rests on our knowledge of the way economic and social mechanisms operate, and the effects of incentives and constraints in different situations. If we are successful in pursuing many policies that increase efficiency, people who lose from one change may gain from many other changes. Since the gains exceed the losses overall, however, we could expect most people, or the average person, to gain in prosperity as the value of overall economic activity increases.

Rule-utilitarianism

There is a version of utilitarianism, called rule-utilitarianism, that may lead to practical rules of action which are similar to those generated by Kant's theory. The rule-utilitarian's position is roughly as follows: Suppose our social goal is to maximise the happiness of society as a whole. If we adopt the act-utilitarian principle ("Each person should act in every instance so as to maximise the happiness of society as a whole"), we encounter coordination and information problems. Following rules of justice (such as those advocated by Kant) may lead to greater long-run happiness for society than would acting in each case according to rules of immediate expediency.

If we had to decide at every traffic intersection whether running the red light would maximise the happiness of society or not, many resources would be wasted trying to answer the question. More fundamentally, the answer depends upon what others decide.

The set of possible actions open to us in most situations is not solely a function of the physical environment, but also depends on the decisions and the thoughts of others. In order to maximise our own happiness, we must know, in part, what others think will maximise their happiness, and what they think will depend upon what we think. Thus, coordination of decisions and beliefs becomes a problem.

Most of the time it is impossible for us to know what would maximise the happiness of society as a whole. Often we are not certain what will maximise our own happiness. Some of the more indirect, but nevertheless significant, consequences of our actions may not be revealed immediately. Acknowledging this, we could not be effective act-utilitarians, even if that were a role to which we aspired.

We might know from many similar experiences in the past, however, that certain choices are likely to maximise utility on average, or in an expected-value sense. A rule-utilitarian might choose actions consistent with the lessons learned from that past experience, even if the salient information at the time a decision is taken suggests a different course of action would maximise utility in this particular instance. The basic idea of rule-utilitarianism is similar to Smith's description of the 'invisible hand' of the market. Smith notes that an individual who "intends only his own gain" is:

> … led by an invisible hand to promote an end which was no part of his intention. Nor is it always the worse for society that it was no part of it. By pursuing his own interest he frequently promotes that of society more effectually than when he really intends to promote it. I have never known much good done by those who affected to trade for the public good. (Smith, 1979: Book IV, 456)

Objections to rule-utilitarianism

Rule-utilitarianism is subject to many of the same criticisms as are made against Kantian theories. Suppose we adopt a good set of rules, including, for example, the rule against lying under oath. We adopt this rule because we believe it will maximise the happiness of society in the long run. What do we do when we encounter situations in which it may not be in society's best interest for a certain person to tell the truth under oath? Do we enforce the rule against lying under oath or not?

If we decide to abandon the rule against lying under oath whenever we can advance the interests of society, we are behaving as act-utilitarians rather than rule-utilitarians. If we refuse in all cases to abandon the rule against lying under oath, we are not really utilitarians at all, because the consequences of obeying the rule do not enter into our decision about whether to obey the rule or not. Thus, the problem arises of creating a rule for when we will abandon the rules. Yet if we had sufficient information to create a rule for when to abandon rules, we would also have sufficient information to choose the best action in any given situation, that is, to be act-utilitarians.

It is possible to conceive of rules that sanction lying under oath in certain circumstances not because those exceptions are guaranteed to yield the best outcome on all occasions, but because they are expected to yield desirable outcomes on average. For example, one cannot be forced to give evidence against one's spouse. Medical practitioners, priests and journalists claim a right to protect the confidentiality of their patients, clients or sources.

A related argument against rule-utilitarianism is that an act-utilitarian can always achieve an outcome that is at least as good as that which is attainable by a rule-utilitarian. An act-utilitarian can always choose to obey the rule, and thus maximise happiness as a rule-utilitarian would. An act-utilitarian has the advantage, however, of being able to abandon a rule when obeying it would not maximise happiness. Thus, act-utilitarianism has all the advantages of rule-utilitarianism, with the added benefit of permitting the maximisation of happiness in cases in which rule-utilitarianism does not maximise happiness (see Smart, 1967 for further discussion).

One problem with the latter argument in favour of act-utilitarianism is that our set of choices depends upon whether everyone is an act-utilitarian or everyone is a rule-utilitarian. The act-utilitarian only gains advantage over the rule-utilitarian if most people are following rules. For example, suppose an act-utilitarian can personally benefit by safely running a red traffic light at some particular time. This action may nevertheless reduce overall happiness, if it undermines other people's resolve to follow the rule of stopping at red lights. If a sufficient number of people stop following the rule, everyone will have to be very cautious at every intersection, in case other act-utilitarians come along seeking to maximise their welfare at the expense of others. Knowledge that others are following a rule will affect the actions, and alter the range of possible actions, for each individual.

Although the act-utilitarian has more choices that the rule-utilitarian, it is not clear that the former is better off than the latter. There are many cases in which we voluntarily restrict our choices in order to maximise our long-run happiness. Recall the story of Odysseus and the Sirens. Odysseus asked his men to tie him to the mast of his ship and not to release him under any circumstances as his boat passed by the island of the Sirens. Odysseus voluntarily limited his set of choices in order to avoid being able to yield to the Sirens' tempting songs. In a similar way, we may voluntarily limit our set of choices, for example, by putting money in a restricted retirement account, in order to avoid spending it when we are young. We limit our options because we believe that will make us better off in the long run.

Similarly, the knowledge that rules constrain others (as well as ourselves) from behaving in an opportunistic manner changes our calculation of the best course of action available to us. Paradoxically, rules that constrain the range of choices, and therefore limit utility maximisation, for particular individuals in particular situations can actually expand the opportunities available to a society as a whole. Such rules can still maximise long-run utility for most people.

The fact that the adoption of rules can expand our set of choices is of fundamental importance to the accumulation of wealth and to economic prosperity. An individual taking out a loan agrees to be compelled to repay the loan with interest, according to a schedule of payments of principal and interest. When the payments on the loan fall due, the borrower would prefer not to make them. If necessary, the courts will enforce compliance with the loan contract. The payments are in an *ex post* sense involuntary, even though they are *ex ante* voluntary. The borrower agrees at the outset to be bound by the terms of the contract because otherwise the loan would be unobtainable. Imposing a constraint on *ex post* behaviour actually expands the *ex ante* possibilities for the borrower.

More generally, in agreeing to live under the rule of law, we give up some degree of freedom in order to have the remaining freedom better protected through cooperative action. The knowledge that the laws will apply to all members of society allows us to predict better the behaviour of others and thus alters the set of choices it is optimal for us to make.

Conversely, it may be optimal for each person to follow the socially optimal rule but only if most people in the society follow it. Suppose, for example, that we have a rule stating that everyone must stand in a queue while waiting for a bus. Each person enters the bus in turn, and the person who arrived at the bus stop first gets on the bus first. Call this the 'bus rule'.

Suppose that most people in society *A* are obeying the bus rule. Now suppose that a large number of immigrants from society *B* come to live in society *A*. There are now twice as many *B* people as there are *A* people.

The *B* people are not familiar with the bus rule. Thus, whenever the *B* people arrive at a bus stop, they crowd in front of the *A* people and force their way on to the buses, violating the bus rule. The *A* people are now very unhappy, while the *B* people are no more unhappy than they were in society *B*, a land which never enjoyed the benefits of the bus rule.

The question now is: should we recommend that people continue to follow the bus rule? Suppose the cost of forcing the *B* people to follow the bus rule exceeds the social gain from doing so. Then the happiness of the *A* people would be maximised by also abandoning the bus rule. For if the *A* people abandon the bus rule, they too may push their way on to the buses and possibly secure seats, while the *B* people are no worse off. The bus rule only maximises happiness if both the *A* and the *B* people voluntarily follow it.

If, as rule-utilitarians, we abandon a good rule (such as the bus rule) just because it is not optimal at the present time, very few good rules will survive. The only surviving rules will be those that are easy to enforce and the observance of which provides immediate and obvious benefits. We argue in Chapter 3 that unrestrained democracy tends to focus on immediate and obvious effects while ignoring costs and benefits that are long term and indirect. It is not optimal to allow a majority of people to do whatever they like.

SUMMARY

We discussed several difficulties inherent in any study of ethics. Ethics has both a theoretical and a practical side. On the one hand, ethical principles would appear by their nature to be eternal and unchangeable. One such truth is the common sense principle: if a person ought to do *X*, then that person can do *X*.

Other maxims of ethics depend for their usefulness on empirical regularities. The degree to which individuals are risk averse or can forgo immediate pleasures for the sake of long-run goals are empirical matters, but they are highly relevant to ethics. Similarly, there is no point in proposing some goal for government policies if that goal cannot be achieved by any actual government. Our ethical theories and proposals must always be tempered by the limitations of human nature and the limitations of our institutions.

Two popular ethical positions were described briefly: the Kantian view and utilitarianism. On the Kantian view, moral truths are eternal, unchangeable, and apply to all rational creatures. Each person, when acting, must will that the rule being followed apply to any person in the same situation. The consequences of a given action are irrelevant to an evaluation of its moral worth. The intentions of the actor are the only factors determining whether an act is right or wrong. Each person must be treated as an end, and never as a means to an end.

The rigidity of Kant's system makes it untenable in ordinary life. There are circumstances in which one must take account of the consequences of an action in order to decide what to do. Consistency is important, but it is not the only important thing in ethics. There are circumstances in which we must use one person as a means to an end, even if this involves violating certain rights of that person.

Utilitarianism has its own problems. On this theory, our goal should be to maximise the happiness of society as a whole. There are many versions of utilitarianism that protect individual rights to a greater or lesser degree, thus mitigating the problem of using some people as a means of increasing the happiness of others. Rule-utilitarianism, in particular, aims to combine the best of the Kantian and the act-utilitarian theories. We noted several conceptual problems with rule-utilitarianism, including the fact that a rule-utilitarian who always follows the rules is not a utilitarian at all. The basic problem remains of deciding when to enforce a rule and when to break it. When do the benefits to society of breaking a rule outweigh the costs in terms of violating an individual's rights, or of eroding general adherence to rules?

We considered problems in resolving conflicts among ethical rules. Some scholars justify social arrangements on the grounds that such arrangements are governed by rules to which any rational person would agree when forming a social contract. Social contract theory urges us to consider which institutions or rules we would consider desirable if we were to abstract from our own particular circumstances and interests. We are asked to consider which rules or institutions are desirable in general, not what happens to further our own interests at the moment.

In practice, we appeal to both Kantian and utilitarian types of theories when attempting to determine the most desirable social and governmental institutions. The question remains, however, what these rules actually are. We take up this question in detail in Chapter 3. We appeal to Kantian arguments when we affirm the importance of individual liberty and horizontal equity. We appeal to utilitarian arguments when we consider measures likely to raise efficiency.

The efficiency criterion, as we defined it and as it is usually applied, treats as equivalent different distributions of welfare with radically different degrees of inequality. A blindness to the distribution of welfare could be seen as both a strength and a weakness of the efficiency criterion.

On the one hand, the efficiency criterion encourages us to focus on rules and institutions, and on longer-term outcomes. Our own particular interests tend to be ephemeral. Maximising our interests at the expense of outcomes on average does not make for good social policy and is the antithesis of ethical behaviour.

On the other hand, a blindness to the distribution of welfare can be criticised because of its indifference to the plight of the poorest and weakest members of society. Some baulk at the suggestion, for example, that it is desirable to implement a policy that benefits many, perhaps even most people, if it is also likely to hurt the poorest among us. Such a policy could, however, increase efficiency as we have defined it.

The efficiency criterion may also focus attention on end states rather than processes or rules. Thus, a policy that promoted efficiency would be less desirable to the extent that it infringed fundamental individual rights or horizontal equity. Individual freedom and the non-discriminatory application of a set of rules or laws are valued for their own sakes and not because they yield more efficient outcomes, if they do.

In this work, we defend the claim that attempting to equalise income or wealth is misguided and is not required in order to have a fair and compassionate society. We argue that government redistribution aimed at obtaining more equal outcomes has severely compromised efficiency, liberty and horizontal equity. Basing welfare policy on compassion, rather than a misguided attempt to achieve equality of outcome, allows a better balance among competing goals.

3

Ethics and Institutions

In this chapter, we discuss the rules and institutions that a rational individual would be likely to include in a social contract. Although we focus on the distribution of the net benefits from economic activity, we also consider the nature of property rights, other legal institutions, and rules concerning the role and powers of government.

In order to decide which rules and institutions are best, we appeal to the Kantian and utilitarian types of ethical theory as discussed in Chapter 2. Our Kantian intuitions may urge us to believe in moral absolutes and to affirm the importance of individual liberty, or the application of a set of rules equally to all. Our utilitarian intuitions may lead us to consider the importance of the average level of welfare. The latter is closely related to the notion of efficiency defined in Chapter 2. If the average level of welfare is increased in a way that changes the distribution of welfare across individuals, those who gain could compensate those who lose and everyone would be better off.

The pursuit of efficiency may yield policy prescriptions that are similar to those that would follow from a belief in the importance of liberty and horizontal equity. We shall see, however, that utilitarian concerns may also lead us to work towards a system of redistribution to assist the poor, even if such redistribution imposes efficiency losses and compromises respect for individual autonomy and property rights. In particular, risk averse individuals may be willing to accept institutions and rules that are likely to produce a lower average level of welfare if in return the undesirable consequences of poverty can be avoided. Empirical judgments about attitudes to risk and average welfare, therefore, will be a key determinant of the rules and institutions likely to be included in a social contract.

Empirical judgments about the way different institutions or rules are likely to operate will also play an important role in our considerations. There is no point in suggesting rules if they cannot possibly be implemented. Similarly, giving governments the right to carry out certain functions may lead to unexpected, and quite undesirable, results if such powers are not used for their intended purpose.

THE VEIL OF IGNORANCE

Rawls (1971) argues that the activities of government ought to include redistribution of income because risk averse people would favour such programmes when forming a social contract. The purpose of thinking about justice in terms of a social contract is that we abstract from the details of our particular circumstances when deciding what an 'ideal' society would look like. The implication is, however, that an individual deciding how to create laws and other social institutions in this "original position" would be behind a "veil of ignorance":

> In justice as fairness the original position of equality corresponds to the state of nature in the traditional theory of the social contract. This original position is not, of course, thought of as an actual historical state of affairs, much less as a primitive condition of culture. It is understood as a purely hypothetical situation characterized so as to lead to a certain conception of justice. Among the essential features of this situation is that no one knows his place in society, his class position or social status, nor does any one know his fortune in the distribution of natural assets and abilities, his intelligence, strength, and the like. I shall even assume that the parties do not know their conceptions of the good or their special psychological propensities. The principles of justice are chosen behind a veil of ignorance. (Rawls, 1971: 12)

Rawls argues that many disagreements concerning the proper arrangement of society result not from different judgments about the desirability of alternative outcomes, or even different assessments of how rules are likely to translate into outcomes. Rather, it is difficult for people to divorce their abstract judgments about the desirability of different rules on average and over a long time horizon from the particular consequences those rules will have in their current circumstances. When judging social rules and arrangements we should pretend that we know nothing about our particular circumstances in life. Ethical behaviour requires us to examine what we ought to do, not what is pleasant for us to do.

Risk aversion and the veil of ignorance

Suppose we accept the position that a good set of social institutions or rules are the ones to which people would agree when first forming a social contract. It is not clear that there is only one rational or correct contract that could be formed.

Rawls observes that a person in the original position, behind the veil of ignorance, cannot rule out the possibility of becoming a member of the worst-off group of people. Rawls asserts that, as a result, a person behind the veil of ignorance would choose a social system that maximises the welfare of the worst-off group. It may be more rational, however, to choose a system that permits greater risk and opportunity in society, even if it does not maximise the lowest level of welfare one could expect to have.

We show in the appendix to this chapter that Rawls effectively assumes that a rational person behind the veil of ignorance is risk averse in the extreme. An individual behind the veil of ignorance who is not infinitely risk averse would exchange a standard of living that is known with certainty (perfect equality) for a risky alternative that also provides a higher average living standard. The intuitive explanation is that the higher average living standard dominates the effect of the increased risk until the risks become large. Average living standards have an effect on welfare that is proportional to the size of the gamble taken. Higher risk, on the other hand, decreases welfare proportionally to the square of the size of the gamble. Thus, a gamble *A* that has pay-offs half as large as gamble *B* will have an average value that also is half as large, but an effect on risk that is only one quarter as large as *B.* For small changes away from certainty, therefore, an increase in average living standards dominates the increased risk unless the degree of risk aversion is infinite.

Even an individual already expecting to be in a risky situation would generally accept at least some degree of any new risk that raises average standards of living. We show in the appendix that this is extremely likely unless the new risk is highly correlated with the existing risky situation and the individual is extremely risk averse.

Empirical evidence on individual behaviour is consistent with another feature of preferences that runs counter to Rawls's focus on extreme risk aversion. The willingness of people to gamble, and the nature of the gambles they find attractive, suggest that most people like distributions of income or wealth that are skewed toward large positive values. Most people are willing to pay more in expected value terms, and also to accept more risk, if in return they get a chance to obtain a very high return. Lotteries, for example, typically offer a few very large prizes, although each ticket has a very small probability of being a winner. Prize money could be separated into many smaller amounts, so that each could be won with a higher probability than the original large prize. The total pool of lottery money would have the same expected prize money and the gamblers would face lower risks. Evidence suggests that gamblers prefer large payments with low probability over smaller payments with a higher probability. That is, they apparently have a preference for positively skewed distributions of prize money.

The demand for positively skewed distributions implies that a rational individual behind the veil of ignorance may opt for rules and institutions that provide opportunities for enjoying high living standards. Conversely, the same feature of preferences implies that people dislike arrangements that give them a substantial chance of obtaining very bad outcomes. Such outcomes could take the form of low incomes, but they could, for example, also take the form of poor health. Buchanan and Prior (1985) discuss the demand for health insurance from behind the veil of ignorance. For simplicity, the following discussion focuses on the case of income fluctuations.

The evidence on behaviour toward risk suggests that a rational individual in the original position would prefer higher average income levels over lower ones, and less variable income distributions over more variable ones. For any degree of variability in incomes, however, the chance of earning large, above-average incomes would be viewed favourably.

A rational, but risk averse, person behind the veil of ignorance would also prefer to reduce the chance of an unfavourable outcome, or attempt to ensure offsetting benefits are available if poverty eventuates. A decrease in wealth, when a person is already poor, will impose losses in welfare that are greater proportionally than the gains in welfare that would result from an equivalent increase in wealth when a person is already well off. A rational, but risk averse, person would be willing to forgo some increase in expected wealth, if the sacrifice allowed the magnitude or probability of low outcomes to be reduced. In short, there would be a demand for insurance against adverse outcomes.

Income transfers are, however, not the only way of insuring against events that force people into poverty. Private financial and insurance markets allow for substantial sharing of risks among people. Such markets also allow people to reduce the costs of variations in income in any one year by spreading the effects over many years.

As we discuss in more detail in Chapter 4, there are also limitations on the extent to which private financial and insurance markets can insure against adverse fluctuations in income or expenditure. In particular, people are never behind a veil of ignorance. When people buy insurance, or deal in financial markets, many determinants of their lifetime incomes are known. A risk averse individual behind the veil of ignorance would like to insure against random variation in many of the factors of which Rawls assumed them to be ignorant, such as "his place in society, his class position or social status … his fortune in the distribution of natural assets and abilities, his intelligence, strength, and the like". Such insurance is unavailable in practice since the outcomes are not random events but known facts about each person's position in life.

We examine redistribution as a form of compulsory insurance in more detail in Chapter 4. We argue that compulsory redistribution implemented through the tax and welfare system is likely to result in large efficiency losses. By comparison, the efficiency losses accompanying the many market and self-insurance opportunities available to individuals will be much lower. In addition, voluntary transfers from families and friends, and support from private charity, are likely to provide substantial resources to those in need, particularly when the government is not itself involved in extensive income redistribution.

Some people may, however, fail to benefit from private charity. As a result, a rational, but risk averse, individual behind the veil of ignorance is likely to desire some form of compulsory public insurance. The very high efficiency costs of taxes and public welfare programmes imply, however, that such a compulsory public insurance scheme is likely to be limited in scope. It will focus on assisting people in dire need. These may be people who have been unlucky in the resources they received from behind the veil of ignorance. They might also, however, have been relatively fortunate in receiving a good set of genes, or wealthy parents, and yet suffered from subsequent bad luck. They may also have suffered from a tragic character defect, or poor information, that prevented them from making the choices that would have yielded the maximum value from their endowments. The reason for a person's current misfortune is irrelevant to our argument as to whether or not that person deserves support. Rather, we claim that a rational individual behind the veil of ignorance is likely to desire transfers to people in dire circumstances regardless of the myriad reasons why someone might end up in such a state.

It is one thing to be concerned about avoiding poverty. It is quite another matter, however, to assume, as Rawls does, that a person behind the veil of ignorance would wish to maximise the value of the worst possible outcome, whether or not that outcome is above a minimally acceptable poverty level. As the expected utility analysis presented in the appendix suggests, we should transfer income to the poor because they desire at least a minimum standard of living, not because they are risk averse.

Utility may become extremely low as consumption approaches a critically lower level. A small addition of resources could make a large difference to someone who is close to that minimum acceptable level of consumption. If so, a rational person would like to be assured of enjoying a standard of living that is above that minimal level. It may be acceptable to endure a substantial reduction in consumption that nevertheless leaves living standards well above the minimum level if, in return, living standards could be guaranteed to exceed the minimum level in all situations. Once guaranteed freedom from poverty, however, the rational person may be content to accept a considerable amount of risk in return for a higher level of expected income.

Maximising the value of the worst possible outcome only makes sense if people are *infinitely* risk averse, an assumption that is inconsistent with human behaviour. Everyone takes risks every day, often in return for relatively small expected benefits. For example, every time one ventures on to the roads one is incurring a non-trivial risk of death, yet we accept that risk in return for quite small expected benefits.

A rational individual behind the veil of ignorance is likely to value efficiency, or higher average outcomes, but also is likely to favour some type of compulsory redistribution as a measure of last resort for the poor. As we shall see later, however, some scholars have suggested a much more ambitious goal for government. They have suggested that equality ought to be something that we value in and of itself, a Kantian-type ideal similar to liberty or horizontal equity. We maintain that such a view is misguided.

OWNERSHIP OF MARKET INCOMES

The goal of equality could be rejected on the grounds that the distribution of income produced by a market economy reflects an appropriate reward for services rendered. For example, the marginalist theory of income distribution is sometimes used to justify a claim that market incomes appropriately reflect the social value of individual contributions to overall economic prosperity. By this argument, people have a right to keep any income they earn through legal means and redistribution, and taxation more generally, is analogous to theft. The argument is similar to the Marxist claim that if the market value of a good or service is determined by its labour content (the so-called labour theory of value) then all income ought to belong to labourers. Any income earned by capitalists is regarded as expropriation.

By contrast, our objection to redistribution is not based on any claim that the income distribution resulting from market activity is just, or inherently superior to any alternative distribution. In essence, the normative basis of our objection to compulsory redistribution is that it compromises the fundamental ethical principles of efficiency, liberty and horizontal equity. The harmful consequences of compulsory redistribution for efficiency, liberty and horizontal equity imply that it can only be justified when it provides significant offsetting benefits measured in terms of furthering these, or other, desirable social goals. We also, however, reject the claim that market incomes reflect the social value of contributions to overall economic prosperity and it will prove useful for later discussions to explain why.

The marginalist theory of income distribution is actually a descriptive, not a normative, theory. It asserts that the market income earned by any person supplying productive resources to a competitive market economy will reflect the *marginal product* or *marginal value* of those resources. Resources include material inputs such as land, animals and minerals; and personal inputs such as time, energy, knowledge and skills. They also include a willingness to postpone consumption until the future or to accept a greater risk of variations in consumption. The first type of input will usually yield income in the form of rents, the second in the form of labour income, and the third in the form of interest receipts, dividends or capital gains on financial investments.

The marginal product of a resource in a particular use is the market value of the additional output that could be produced by employing a small additional amount of that resource, while holding all other resource inputs constant. If resource owners aim to maximise the market income they receive, we are likely to find resources allocated to their most productive uses. If a resource is earning less than its marginal product, a competing employer will have an incentive to offer a higher payment to gain access to the resource. By splitting the difference between the current payment to the resource owner and the marginal product of the resource, a new employer and the resource owner could both be made better off by a new allocation of resources. The process would continue until incomes are bid up to a level that reflects marginal products.

There are many objections to the claim that market wage rates, rents or yields on investments reflect marginal social values of resources. First, the marginal value of services that people supply may be reflected neither in market prices nor in money incomes. For example, many people would consider the services that they voluntarily provide to their family and friends to be among the most valuable activities in which they engage, yet none of these are reflected in their own money incomes.

Even activities that could, in principle, generate a high money income may fail to do so in practice. No one would deny, for example, that the marginal value of the services provided to society by Mozart or Van Gogh dwarfs the personal money income these individuals earned. The full value of their labours was not appreciated at the time, since it is not easy to predict the music or works of art that will be highly valued by future generations. The expected marginal value of new art or music compositions might be reflected in the average incomes paid to artists or composers, but there is great uncertainty about how art or music will be regarded in the future. Similarly, it is costly to discover and arrange more productive uses for resources. The market incomes earned by two identical resources therefore could deviate from each other by the search and other costs involved in achieving a superior arrangement. Market incomes will thus reflect many random influences, or what is popularly termed good or bad luck.

Some other valuable services may not generate high money incomes because the necessary property rights have not been recognised. For example, basic scientific research can pay enormous dividends, but often only a tiny fraction of these benefits accrue to the person who made the discovery. Furthermore, while patents provide some income to their owners, they are of limited duration and can often be circumvented by new inventions.

A second reason money incomes may not reflect marginal social values is that individuals care about more than the market income they earn. Jobs with high non-pecuniary benefits will, however, often have lower wage rates. For example, people care about working conditions, where they live, and how much time they have to spend with family members or friends, or engage in leisure activities or hobbies. They will be happy to accept a job with a lower wage if other benefits more than compensate for the difference in money income. As an extreme example, people who volunteer time to charitable organisations typically receive little, if any, monetary compensation. Presumably, they feel rewarded in other ways, for example, through enhanced self-esteem, or the emotional benefits of the friendships they form.

A further critical assumption underlying the marginalist theory is that the economy is competitive. When there are few good substitutes for productive resources or jobs, either the seller or the buyer of productive inputs will have monopoly power. A particular match between a seller and a buyer will produce rents, measured by the difference between the total value produced by this allocation of resources and the total value produced by the next best alternative allocation. The allocation of such rents between the buyer and the seller of a resource may be indeterminate.

A related issue is that the marginalist theory of income distribution implicitly assumes that input proportions can be varied continuously. If instead resources can only be used in fixed proportions, or can only be added in large lumps, it may be difficult to identify individual marginal products. In some of these cases, however, we might be able to extend the theory by examining the returns to realistic increments in resource use, or by extending the time horizon to allow greater flexibility in using alternative production technologies. Moreover, wage and rental rates will reflect the contribution a resource could make to productivity in a wide range of alternative industries and not simply the effect of marginally expanding its current use in any one industry or activity.

The marginalist theory also implicitly assumes that production processes are characterised by constant returns to scale. A uniform expansion in the use of all inputs will lead to an expansion of output by the same proportion. One can show that paying all resource owners according to their marginal products will under these conditions exactly exhaust the total value of output. If a production process is not characterised by constant returns to scale, however, paying resource owners according to their marginal products will either be infeasible (if there are increasing returns to scale) or not exhaust total income (if there are decreasing returns to scale).

Suppose that we put aside such objections and accept the claim that market incomes paid to resource owners approximate the marginal products of those resources. The marginalist theory still could not account for the incomes that each individual enjoys. Market incomes earned by individuals reflect the ownership of different resources in addition to the marginal products of resources. Some resources, such as skills, work experience or knowledge and physical or financial capital, are accumulated through prior investment by the individual (or the individual's benefactors), and thus could reflect past productivities. But genetic endowments, in particular, are inherited, and it is beyond the domain of economics to explain the distribution of these gifts across the population.

There is, however, a more fundamental objection to using the marginalist theory to assert that the distribution of income arising from market activity is just, and tampering with it is unjust. Such an assertion would be equivalent to claiming that people ought to receive, as a matter of justice or fairness, what economics implies they are likely to receive in a competitive market economy. Yet no ethical premise has been specified that would justify such an inference. By the same token, one cannot conclude that the distribution of incomes produced by the market is unethical or unjust.

The marginalist theory is an attempt to explain why incomes are what they are. In order to discuss what ought, or ought not, be done about inequality one needs to go beyond descriptive theories and discuss explicitly the normative principles that ought to guide individual or social action. Our objection to redistributing income among people, each of whom has a decent standard of living, is based on the claim that the benefits of such redistribution are insufficient to compensate for the infringements of individual liberty and the efficiency costs that would be incurred.

Ownership and efficiency

Hayek argues that it is not necessary to justify the distribution of income produced by the market:

> There is no need morally to justify specific distributions (of income or wealth) which have not been brought about deliberately but are the outcome of a game that is played because it improves the chances of all. In such a game nobody 'treats' people differently and it is entirely consistent with respecting all people equally that the outcome of the game for different people is very different. It would also be as much [of] a gamble what the effects of any one man's efforts would be worth if they were directed by a planning authority, only that not his knowledge but that of the authority would be used in determining the success or failure of his efforts. (Hayek, 1976: 117)

We can examine whether institutions, or rules of the game, are just or likely to yield more or less desirable outcomes. The specific outcomes, however, are a product not only of the rules of the game, or the institutional arrangements, but also of many particular facts and random occurrences. No one person or organisation has any control over all these facts or events, so it makes no sense to hold anyone responsible for their existence.

What we can attempt to do, however, is to examine different rules or institutions and decide which of them have in the past yielded, and are more likely in the future to yield, desirable outcomes on average and over the long run, that is, increase efficiency. Thus, Epstein (1995: 56–57) argues that self-ownership of labour time is efficient:

> … any system that begins with the collective ownership and control of individual talents builds a form of institutional paralysis into the very fabric of social lives. So long as *A* does not own *A*'s labour … *A* is no longer entitled to sell *A*'s labour, and therefore has less of an incentive to improve it. Likewise things that are made by his hands are no longer his either, but instead belong (if they are made at all) to that group which owns the labour in question. Finding those individuals or groups who are co-owners in *A*'s labour is no easy task …
>
> Even though talent, circumstance, and luck play a role in human behaviour, we are all spared an enormous administrative burden if we mutually renounce any claim to these assets of others. A rule of self-ownership, far better than any of its alternatives, allows us to move on with the business of life. A rule of self-

> ownership selects the *single* person to be the owner of each person's natural talent, and picks that person who in the vast majority of cases tends to value those assets the most: each obtains control over his or her own body. At least for adults (and there are, of course, qualifications for children), the rule offers the shortest path from initial entitlement to productive human activity.

Similarly, Epstein (1995: 59–63) provides an efficiency justification for the common law rules of "natural occupation", or "first possession" for land or "the rule of capture" for wild animals and material things:

> Like the autonomy rule, the first-possession rule imposes costs on those who are excluded. But also like the autonomy rule, it provides compensating benefits – most notably, the security of possession – that justify its place in the social order. The rule allows for decentralised control of material resources at a low overall cost. Those who are unable to acquire goods by first possession are able to acquire them by purchase financed by money, goods, or land that they acquire from the use of their own labour. Since two individuals can make gainful exchanges without the participation of the rest of the world, the overall size of the gain is so large that we need not trouble ourselves over its distribution. Quite simply, the gains from the use or exchange of a very broad class of resources are higher than the losses others bear from exclusion. (Epstein, 1995: 62)

Epstein (1995) goes on to consider the problems that arise when property is unlawfully taken from someone who first possessed it. If the original claimant acts quickly enough, and the courts respect the right of first possession, the original owner could recover the property. If the new owner has used the property for some time, however, it might be efficient to allocate the property rights to that new owner:

> … in practice all legal systems have adopted statutes of limitations to cope with the problem, so that if a suit is not brought within some fixed number of years the property is lost … Some rule of this sort is essential, first because it protects those titles that are valid from false suits brought by others, and second because it again creates a single owner needed to bring property back into the stream of commerce: who will purchase or improve land when someone else may be able to snatch it away? …
>
> The second major challenge to a system of first possession arises when neither claimant has the best claim to the disputed property. Thus if *C* takes property from *B*, which *B* has taken from *A*, *A* may surely recover the property from either *B* or *C* … [but] if *A* has left the scene, what are the relative positions of *B* and *C*, when neither is the first possessor of the property? …
>
> To meet this difficulty, courts uniformly hold that *C* cannot defend her possession against *B* by showing the better right of the third party, *A* (the so-called *ius tertii*), in any suit brought by *B*. Under this system, title becomes a relative notion, so that prior in time becomes higher in right, with first in time

> being just a special case of a general principle. With the passage of title, all higher claimants will either bring their suit or have their action barred by the passage of time, so that we are once again left with a single person vested with all the attributes of a common law owner. (Epstein, 1995: 64–66)

For Epstein, the major social gain from clear and enforceable property rights is that only then can we develop commerce and exploit the gains from trade and specialisation:

> When property rights become well defined, non-owners (who still own their labour) receive enhanced opportunities to gain from exchange. What is lost to latecomers from the world of acquisition is provided for in the world of trade and commerce for the betterment of those who did not acquire anything from the original commons. (Epstein, 1985: 11)

Thus, as an implication of the desirability of efficiency as a social goal, we conclude that the person who owns a particular resource ought to be the primary claimant on income derived from its use. This argument allows for the possibility, however, that individuals could voluntarily agree *ex ante* to be coerced into contributing *ex post* to the supply of certain types of jointly used facilities or services or to alleviate poverty. It may be efficient to supply public goods or to take joint action to ameliorate certain externalities, or desirable to transfer income to the poor for the reasons discussed above. For these purposes, taxation of market income may be required. We thus are led to consider the likely consequences of different social decision-making rules or institutions, and hence the type of joint actions that may be considered desirable.

SOCIAL DECISION-MAKING RULES

One can use the construct of a rational individual behind the veil of ignorance to decide which type of social decision-making rules or institutions are most desirable. We argued in Chapter 2 for the rationality of a democratic type of arrangement in which people are given an equal say in determining social goals. Such an arrangement does not, however, uniquely determine the exact decision-making rule.

Voting rules

The classic treatment of the determination of voting rules from behind the veil of ignorance is Buchanan and Tullock (1965). They suggest that the rational individual behind a veil of ignorance would maximise the expected net gains from collective action, or efficiency, but the costs of making a decision are taken into account.

We might identify a theoretical *first best* allocation of resources as a Pareto optimal one (perhaps randomly chosen from among many) given everyone's endowment of resources, tastes and so on and assuming there are no decision-making costs. We can then measure actual outcomes as costs relative to this unobtainable imaginary first best allocation. The costs so identified take two forms. One is the departure from the Pareto optimal outcome. The other is the resources consumed in the decision-making process. One can compare this analysis with Epstein's analysis of the law:

> The social question, however, is concerned with the extent of the fall from grace … Within this framework, the social function of the law is to minimize the sum of the administrative (including error) costs and the costs associated with the creation of poor incentives for individual action. (Epstein, 1995: 32)

Decision-making costs

Buchanan and Tullock (1965: 68–69) suggest that decision-making costs would generally be expected to rise with the amount of agreement one needs before making a decision. At one extreme, unanimity is likely to lead to very large negotiation costs because of what we termed above the 'hold out' problem:

> With … unanimity … each voter is a necessary party to any agreement. Since each voter, then, has a monopoly of an essential resource (that is, his consent), each person can aim at obtaining the entire benefit of agreement for himself.

For decision-making groups of smaller size:

> … there is apt to be little real bargaining. If one member of a potential agreement asks for exorbitant terms, the other members will simply turn to someone else.

At the other extreme from unanimity, Buchanan and Tullock assume that a dictator, chosen at random from the population, would take no time or resources to decide what to do. People other than the dictator, may, however, expend resources attempting to influence the choice (as they did, for example, when monarchs had substantial real power). As the proportion of the population needed to assent to a change rises, the resources expended to influence the governing group are likely to fall (since the spoils need to be shared among a larger group), but the costs of reaching a consensus within the group are likely to rise. The net result may be a decision-making cost curve similar to the one drawn in Figure 3.1.

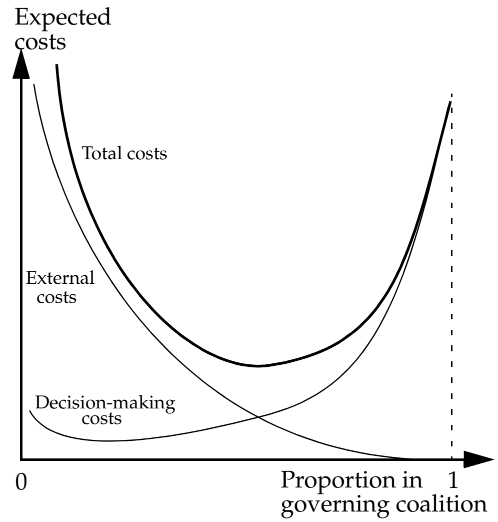

FIGURE 3.1: Expected costs relative to 'first best'

External costs

The other cost curve in Figure 3.1 is labelled as the external costs. It represents the expected costs that an individual who is outside the governing coalition might bear.

In a democratic society, government is the only organisation with a legal right to coerce others. By making the right to coerce belong to a monopoly, it may be possible to avoid struggles among powerful competing groups.

The ability to coerce gives government a unique advantage in some respects. Other organisations or institutions can exploit individual effort only after voluntary consent. Because governments can legally coerce people to act against their will, however, governments have the potential to overcome so-called 'free riding'. Everyone who benefits from a jointly provided service (or public good), such as law enforcement or national defence, can be forced to contribute to the cost of provision. A situation of chronic under-supply, as everyone waits for neighbours to contribute first, might thereby be avoided.

Individuals behind the veil of ignorance would also recognise, however, that they could value public goods at less than the amount they are forced to contribute. A major disadvantage of forcing people to contribute to the cost of providing services, or divorcing payment for services from the decision to use them, is that information about the value people place on the services is lost. When a person pays voluntarily for a good or service, the amount bought equates the marginal benefit to the price paid. If the person bought more than this amount, the marginal benefit of the last unit would fall short of the price, and the person would be better off reducing purchases. If an individual bought less than the amount that equates the marginal benefit to the price, buying more would make that individual better off. Each buyer thus has an incentive to reveal what would otherwise be the personal and private marginal value of additional supply of a product. So long as each individual acts rationally, purchases will be adjusted until all marginal benefits equal the publicly known purchase price. One can regard the entire system of voluntary transactions as an information-revelation mechanism that gives each person an incentive to reveal honestly what would otherwise be private marginal valuations. When products are provided through coercion, however, this source of information on individual valuations is lost.

Furthermore, once an apparatus is in place to force contributions for jointly chosen activities, such a mechanism can be abused. If unanimity is required, the society could implement only pure Pareto improving policies. Since everyone has to agree, a policy would be unacceptable if it made anyone worse off. When unanimity is not required, however, the governing group could force contributions that provide private benefits to itself. For example, if just one person were able to make decisions for the group as a whole, most people exercising that perogative would make themselves very well off at the expense of others. More generally, when the decision-making group is larger, an individual is more likely to be part of the governing coalition and therefore less likely to bear external costs.

Overall costs will be minimised by choosing a governing coalition that falls short of unanimity, and may approximate a simple majority. The decision-making rule is, however, likely to depend on the nature of the decision. In particular, constitutional questions that pertain to 'rules of the game' may potentially impose far greater costs on people outside the governing coalition. We might expect therefore, and indeed typically do find, that such decisions are made by 'super-majorities' of one sort or another. For example, to become law, a constitutional amendment in Australia needs the support of a majority of voters overall and also a majority of voters in a majority of states.

Representative democracy

The above analysis concerned decision-making on particular policies using a direct democracy mechanism. For example, we could imagine such a mechanism being a referendum either on a particular issue or constitutional rules of one sort or another. It would be costly, however, to make all government decisions through direct voting by each citizen. In practice, most decisions are delegated to elected representatives.

Elections and majority voting as a constraint

Delegating a monopoly right to coerce to a small number of decision-makers opens up the possibility that large external costs could be imposed on particular individuals. One mechanism for limiting governmental abuse of power is periodic elections involving competing coalitions. There are a number of reasons, however, to expect this constraint to be weak in practice.

Since elections are infrequent, and voters have short memories, governments can ignore voter sentiment for long periods of time. Elections involve only a small number of candidates, resulting in a limited range of policies being made available to the voters. Most voters, therefore, find it difficult to match their own preferred outcomes with the policies on offer.

Conversely, politicians may have difficulty discerning what the majority prefers. Choosing policy packages by majority voting can lead to a continual cycling among preferred policies. For example, suppose there are three policies, A, B and C but, because of their costs, a majority of voters will support at most one of these. Suppose one third of the voters favour each of the rankings (where the symbol $\succ$ means prefer): $A \succ B \succ C$, $B \succ C \succ A$ and $C \succ A \succ B$. If A is the current policy, a proposal to replace it by C will garner the support of the last two groups of voters. If B is the current policy, a proposal to replace it by A would be supported by the first and third groups of voters. Finally, if C is the current policy, the first and second groups would support a proposal to replace it by B instead. No matter what policy is being implemented, a new majority can be found that favours something different.

Information in the political market place

As noted above, since the political system is based on coercion, it is likely to be much less efficient at revealing information than a system of voluntary transactions. In addition, a voter has a weak incentive to become informed, because making a more informed choice is unlikely to affect individual or family welfare. It is extremely rare for an election to be decided by just one vote. The expected value of making a more informed voting decision would fall short of even trivial costs of gathering the relevant information.

Voting is often seen as an expressive act, similar to cheering for a football team while watching television. Such support for the team cannot possibly make a difference to the outcome of the game, yet people enjoy participating and taking sides. One can often make sense of television news and current affairs shows, newspapers and so on when these are seen as part of the entertainment industry, rather than a genuine attempt to convey useful information or to promote intelligent debate on issues.

Even if voters had an incentive to become well informed before making choices, information about candidate or party performance is difficult to obtain. Many voters (and journalists) economise on information costs by judging politicians on their intentions rather than their performance. Voters are interested to know what motivates political decisions, not details about the effectiveness of different policies, the former being much easier to discover than the latter. Thus the media often focus on the ethics and values of politicians, rather than on the way politicians vote (or even whether they vote) on legislation.

Psychologists and political scientists have examined the way perceived character traits in candidates such as integrity, empathy, effectiveness, competence and leadership affect voter choices. There is evidence that all of these factors matter, although by different degrees for different voters. Sometimes, the same character trait might be seen in both a positive and a negative light. For example, integrity might be seen as inflexibility, empathy as insincerity, and leadership as despotism. Candidates attempt to portray their opponents' characteristics in a negative rather than a positive light.

In so far as there is any interest in the effects of policies, both voters and politicians tend to focus on immediate, obvious and dramatic effects. The unintended and gradual effects are often ignored, even if these are, in the long term, of greater consequence. In particular, electoral cycles need to be short if elections are to be effective in forcing politicians to pay attention to the demands of voters. On the other hand, short electoral cycles encourage politicians to focus on short-term policies.

Democracies tend to cater to vested interests at the expense of the general public interest. If a voter is going to spend time and other resources monitoring government policies, lobbying politicians, or financing studies or political campaigns, the resources will have a greater return if the monitoring effort is focused on the voter's special interests.

Anyone who has lived in a democracy can think of ways in which governments cater to vested interests. The tendency of governments to assist vested interests has, however, been documented in a more systematic way by research in economics and political science. Representatives from electorates with substantial industry that competes with foreign imports typically favour more tariff protection, despite efficiency losses for the community as a whole. Agricultural policy encourages the production of food that is destroyed or exported under subsidies raised at considerable expense to taxpayers.

Rent controls appear to favour renters at the expense of landlords, but ultimately lead to a deteriorating housing stock and to the destruction of entire neighbourhoods as crumbling buildings are flanked by deteriorating public spaces around them. Public school teachers lobby for educational reforms that increase the number of teachers, make it more difficult for new people to become qualified as teachers, reduce teacher workloads, and reduce the skills and knowledge learned by students.

Doctors and other professionals lobby for licensing restrictions that reduce competition in their industries and impose costs on the rest of society that exceed the benefits they obtain. Legalised monopolies in product markets, legislative support for union monopolies, and public ownership of potentially competitive enterprises similarly provide substantial benefits for a few at the expense of the many. The list of such policies is long.

Politicians can gain, and retain, office by putting together a package of policies that focus on numerous special interests, even if such policies compromise efficiency, or the general public interest. Since information about the ultimate effects of policies is difficult to obtain, programmes aimed at assisting different special interest groups often conflict with each other. For example, tariffs favour industries that compete with imports and harm exporting industries (among others) while export subsidies have the opposite effect. A politician proposing both policies may, however, earn the support of vested interests from both sectors of the economy. Each group focuses on the beneficial effects of the policy aimed at its sector, while ignoring the harmful effects of the policies aimed at furthering competing special interests.

High information costs in the political market place also allow politicians to hide the true intention of policies. For example, regulatory controls are often aimed at producing outcomes of interest to special groups, with the costs hidden more effectively than would be the case with direct transfers out of general tax revenue.

Information costs have a number of implications for policies ostensibly aimed at assisting the poor. Voters who do not have a vested interest in the issue are likely to be satisfied by the thought that money is being devoted to the problem. Some voters may be suspicious that the problems appear to be intractable despite billions of dollars allocated to their eradication. Even so, if citizens spend time investigating government activities, their labours are likely to be better rewarded if they focus on issues closer to their own self-interests.

Voters, and civil servants, who have a vested interest in the alleviation of poverty may be more concerned with how policies affect their own power or livelihood than with what they achieve for the ostensible targets. In particular, it would be unfortunate for those with vested interests if effective policies were to be found and applied, and the reason for the existence of the programmes were to disappear.

Public versus private interest in a democracy

Despite the tendency for democratic politics to pander to vested interests at the expense of the general public interest, there is also a limit to this process. For example, suppose an *X* Gift Bill were proposed that would tax each person in New Zealand an extra 5 cents per year. Suppose that all the proceeds from the tax would be automatically deposited in *X*'s bank account. *X* has no special needs or merits that might justify the huge gift from the taxpayers. *X* is just a lucky person who hopes to benefit from parliament's favours.

The loss to each person paying the 5 cents would be very small, and the gain to *X* would be very large. Yet the extremely low cost of comprehending the effect of the policy would make it impossible to pass by majority vote. A proposal needs to have at least the appearance of furthering a wider public interest to gain majority support.

Political lobbyists emphasise the ostensible public benefits of policies. The private benefits provided to vested interests are rarely mentioned. It is more acceptable to claim that tariffs are desirable because they "protect our country against unfair foreign competition" than to assert that tariffs are desirable because they increase the value of an individual's share portfolio or wages.

Shared community values or ethical standards are one set of wider public interests that are often acceptable as a basis of support for policy positions. The high cost of becoming informed about the consequences of policies makes ethical standards, and the expressed intentions of politicians, more important to voters and the media. Ethical positions therefore are relevant to political discourse because they help determine the type of policies that can gain majority support.

Ethical positions are, however, not the only widely held community beliefs to which politicians can appeal. Attitudes, beliefs or actions that are popular are not necessarily right or moral. Politicians often appeal to xenophobia, racism, religious fanaticism and many other beliefs (such as the view that governments control economic growth rates) that, although widely held, may nevertheless be erroneous. For example, the notion that tariffs are ultimately paid by foreign producers, rather than by other domestic producers and consumers, makes tariffs attractive to people motivated by xenophobia.

Politics is, unfortunately, far removed from rational discussions of ethical principles. In particular, nothing inherent in democratic political institutions can guarantee that the only beliefs or values influencing political outcomes are those that would be considered ethical or moral by an objective observer. Nevertheless, public discussion of ethics, along with ideas and evidence on the effects of policies, no doubt play a role in influencing political outcomes.

Limitations of bureaucracy

Suppose we ignore the problems considered above and assume that representative democracy is effective at forcing politicians to implement only those policies that would be considered ethical by a majority of voters from behind the veil of ignorance. Even if this were the case, politicians may not be able to implement the desired policies.

The government, or responsible minister, relies upon a bureaucracy for information and advice, while also depending on the same individuals to carry out policy directives. Bureaucrats have their own goals and interests, and respond to incentives and constraints, as everyone does. By withholding information, or otherwise distorting the information they provide to the politicians, bureaucrats can pursue their own self-interest or preferences.

Even when bureaucrats are attempting to pursue what they see as the social interest, there are systematic tendencies that can drive a wedge between perception and reality. There is a natural reluctance to admit to past mistakes. Failed government programmes inevitably lead to calls for more intervention rather than an admission of failure and return to the *status quo ante*. There is also a natural tendency to seek solutions to perceived problems that involve more activity by the bureaucracy, rather than allowing individuals to act on their own behalf.

The superior performance of newly privatised firms in New Zealand, the United Kingdom, Australia and elsewhere relative to the government enterprises that they replaced (see the references cited in Chapter 2) illustrates the limitations of bureaucratic management. The inefficiency of publicly owned enterprises results in part from a lack of accountability of the managers of such enterprises relative to their private sector counterparts, as has been demonstrated by corporatisation. Under corporatisation, a board of directors is appointed and made responsible for achieving more transparent and focused objectives, managerial rewards are closely related to achievement of the stated objectives, and political directives must be explicit and open to public scrutiny. While corporatisation of publicly owned enterprises has increased efficiency, subsequent privatisation has delivered additional efficiency gains.

The tendency of representative democracy to favour vested interests at the expense of the general public interest is another source of inefficiency in publicly owned enterprises. Public enterprises typically provide higher wages, better working conditions and higher levels of employment than those provided by private-sector counterparts. Some consumer groups, such as those living in remote areas who are more expensive to serve, also benefit from cross-subsidies, or prices that are below the costs of service. The redistributions facilitated by public ownership are inefficient, however, since the benefits delivered to the favoured groups are less than the costs imposed on taxpayers and consumers. Taxpayers are called upon to finance investments at uncompetitive rates of return, while consumers who are less expensive to serve, or who have relatively inelastic demands, are often charged much more than the costs of serving them. Hartley and Trengove (1984) documented these tendencies in the case of the former publicly owned State Electricity Commission of Victoria, while Hartley and Trengove (1986) discuss a number of factors that determine the extent of benefits accruing to employees, or politically favoured customers, of publicly owned enterprises.

Limiting abuses of representative democracy

The many defects of representative democracy as a social decision-making mechanism imply that a rational individual behind the veil of ignorance would strictly limit the power of elected governments. Limitations can take several forms. Powers may be divided between the legislative, judicial and executive branches of government. The right of government to make new laws may be constrained by formal or informal constitutions, or by strong social norms. Finally, competition between governments can limit the freedom of action of each government.

Constitutional decision-making

Buchanan (1975) discusses constitutional constraints on government from the perspective of social contract theory. He asks what type of restrictions free individuals voluntarily forming a society would place on government:

> The terms must include, first of all, some statement of limits on the behaviour of any person with respect to the positions of other persons in the community … As they enter genuine society from anarchy, persons … accept rules governing their own behaviour in exchange for the like acceptance of such rules on the part of others. Secondly, the basic contract must define the positive rights of possession or domain over stocks of goods, or more generally, over resource endowments capable of producing final goods … [T]he inclusive constitutional contract must also make explicit the terms and conditions of enforcement … Finally, the basic contract must define the rules under which the collectivity must operate in making and in implementing decisions concerning the provision and financing of 'public goods'. This set of terms will specify in detail the operation and the limits of the productive state, the legislative aspect of collective organization. The rules and institutions of this productive state may, in themselves, incorporate several dimensions. The contract should indicate the allowable range over which collective action may take place. That is to say, some restrictions on the type of goods to be provided and financed collectively must be included … Within these defined limits, allowable departures from unanimity in reaching collective decisions should be specified.

In practice, a written constitution, or set of inherited laws and conventions, often constrains the activities that a government is permitted to undertake. These laws or conventions typically are interpreted by a legal authority that is largely independent of the current government. While the government can change laws, its actions are also subject to a rule of law.

Constitutional constraints protect individual rights and limit the tyranny of the majority, or even the tyranny of the decisive marginal voter or party. As Epstein (1993: 78) observes:

> In some markets the government has a high degree of monopoly power. It may be the only party that can operate the public roads, issue building permits, or allow firms to do business in corporate form. Unlike the private monopolist, its power cannot be eroded by the entry of new firms, but is perpetuated by a legal prohibition against entry by new rivals. The major task of constitutionalism, therefore, has been to forge a system in which these government excesses are curtailed while leaving the state sufficient power to discharge the necessary tasks of governance. A system of unrestrained political power does a poor job in setting the right balance.

For example, Epstein (1985, 1993) argues that the takings clause of the United States Constitution: "Nor shall private property be taken for public use, without just compensation", might be seen as a brake on some of the excesses of majority rule:

> By forcing the winners to fully compensate the losers for the deprivation of their rights, the losses that the winning coalition could otherwise externalise on losers are brought back to bear on them … The compensation requirement preserves a sense of fairness across parties that is captured in the Paretian formula. It also changes the mix of legislative schemes that are generated by altering the incentives to all the players within the system and hence the selection of projects that go forward and those that do not … By weeding out poor legislative schemes, these allocative consequences are always positive and, in most instances (if historical experience is a loose guide), seem to justify the substantial administrative burdens they impose. (Epstein, 1993: 85)

Similarly, he remarks that the basic plan of the United States Constitution is to "limit government by indirection":

> At every turn the constitutional concern is with preventing the concentration of power in a few hands. The complex provisions on selecting public officials are designed to control the risks of popular sovereignty by making sure that the electorate cannot speak at once and with one voice …
>
> Similarly, the jurisdictional limitation on federal power is a further effort to limit the power of the sovereign by ruling certain areas out of bounds to collective governance. Federal powers are not plenary but enumerated, and confined largely to matters of Lockean public good … The system of checks and balances, or separation of powers, then makes it more difficult to exercise the powers granted to government … By setting strength against strength, the system helped to guarantee the liberties of the individual. So long as the temptation to abuse power exists, power must be divided in order to be checked. (Epstein, 1985: 16–17)

Since constitutions are in practice determined through a political process it might seem paradoxical that they would ever be devised. Why would a super-majority constrain the types of decisions a majority can later make? As we noted in our discussion of rule utilitarianism, a rule expands the possible actions that may be taken today by limiting subsequent freedom of action. What appears to be a constraint on the range of possible actions by a majority may in fact expand the set of opportunities available to them. If individuals had no confidence that the government would, or could, enforce the law, the benefits of forming a government in the first place would be lost. Citizens would behave as if they lived in a state of anarchy.

A constitutional constraint would be of no value if it could be easily evaded. On the other hand, constitutional rules could become more harmful than beneficial if there were no way of adapting them to changing circumstances. The usual compromise is to make alteration of constitutions possible, but difficult. Typically, much more than a simple majority in a single legislature is required.

For example, Article V of the United States Constitution allows amendments to the Constitution to be proposed by a two-thirds vote of both houses of Congress or by a convention called by Congress on the application of the legislatures of two-thirds of the states. Amendments that have been proposed by Congress must be ratified by three-fourths of the state legislatures or by conventions in as many states. Congress decides which method will be used and sets the time limit for ratification.

Section 128 of the Australian Constitution requires a proposed amendment to be passed by a majority in both Houses of Parliament or, if it is passed by only one house, to be passed twice within a three month period. In either case, the proposed amendment must then be presented to the voters in a referendum. Before it becomes law, the amendment has to be approved by a majority of voters in a majority of the Australian states, as well as a majority of voters overall.

These rules could be avoided by a revolution that makes the whole constitution invalid. It is more difficult to mount a revolution, however, than it is to meet the requirements for amending the constitution.

In New Zealand, as in the United Kingdom, there is no written constitution. However, the powers of the government are derived from the common law powers of the Crown. As with other common laws, these implicit restrictions on the exercise of government power can change only gradually and as a result of new precedents being established by a judicial review process that is independent of government.

The New Zealand and United Kingdom approach may be less different from the American and Australian system than they first appear. The written constitutions in the latter countries must also be interpreted by the courts and the interpretations evolve over time, as a result of legal cases that set new precedents. Some of these interpretations appear to depart considerably from a more literal interpretation of the original provisions. The written texts may do little more than slow down the rate of evolution of judicial interpretation of constitutional law.

Because constitutional provisions are expected to persist for a longer period of time than ordinary laws, individual vested interests become harder to determine. A provision that favours a particular family now may work against that family's interests in only a generation or two. In contrast to the dominant role of vested interest in ordinary law-making, the average or long-run consequences of a rule are more likely to influence views on proposed constitutional constraints. People may even reject proposed constitutional rules that are beneficial to someone in their particular circumstances but that are harmful on average.

The knowledge that constitutional rules are likely to be hard to change in the future also raises the costs of making a bad decision. When creating constitutional laws, therefore, there may be a tendency to favour provisions that further the general public interest rather than the individual special interests of an ephemeral majority.

The legal system

Although the electoral mechanism places only weak constraints on political behaviour, other controls may be more restrictive. In particular, successful democracies are characterised by a separation of powers between the legislative and judicial branches of government. (In the presidential system, the legislature and the executive are also more independent of one another than they are in a parliamentary system.) Raising taxes to finance jointly used goods or services is a principal responsibility of the legislature. The legislature can enact new laws not only to support its fiscal responsibilities but also to regulate social interactions of many different kinds. An independent set of institutions is responsible for enforcing the law, including placing written or common law constitutional constraints on the actions of the legislature.

In a common law system, the judicial branch can change the law by establishing new precedents. These changes are, however, more evolutionary in nature, and are made in a more decentralised way, than are the changes enacted by the legislature. Even in legal systems in which there is no formal appeal to precedents, courts often rely on past cases to decide how new laws ought to be interpreted and applied.

The major advantage of a decentralised system for applying laws and setting new precedents is that, in many cases, the inefficiency of an old rule will become apparent only as it is applied in particular contexts. It will be impossible to predict the precise implications of a particular rule in all the situations in which it will apply. Only the general tendencies of rules will be apparent, and this may result in a greater focus on efficiency:

> [I]n the usual sense of purpose, namely the anticipation of a particular, foreseeable event, the law indeed does not serve any purpose but countless different purposes of different individuals. It provides only the means for a large number of different purposes that as a whole are not known to anybody. In the ordinary sense of purpose law is therefore not a means to any purpose, but merely a condition for the successful pursuit of most purposes … It certainly has … developed because it made people who operated under it more effective in the pursuit of their purposes. (Hayek, 1973: 112–113)

A bias in the cases that are relitigated may be the major systematic factor prompting the legal system to evolve toward an efficient allocation of rights and responsibilities. If a legal rule is inefficient, the affected individuals could rearrange rights and obligations so that the gains to one party more than compensate for the losses to the other party. An efficient legal rule, on the other hand, can only be overturned by redistributing from one party to the other. There are no net gains that could be used for compensation. This is simply an implication of the definition of efficiency. The result is, however, that inefficient legal rules may be more likely to be challenged than efficient ones. Cases will be brought forward to have the old rule overturned and a new precedent established.

It is possible, however, to overstate the effectiveness of selective relitigation in eliminating inefficient laws. Legal cases are decided on the basis of both factual and legal issues and it is often unclear whether a prevailing legal standard is inefficient or not. Furthermore, higher courts may overturn judgments of lower courts for reasons other than efficiency, and there is no systematic mechanism encouraging the courts to countenance efficient rules as opposed to rules that satisfy some other criterion.

In particular, judgments must be seen as just. For example, suppose we could convince everyone that a homeless person with no surviving family members had recently committed a murder, although the person was innocent. Imposing a severe penalty on such a person might deter others contemplating murder. The gain to society from punishing the innocent person could far outweigh the costs. It might therefore be efficient to proceed with such punishment, but it would not be just.

As with political decision-making, widely held ethical and other beliefs help delineate what is considered just. Furthermore, factors other than ethical beliefs determine the outcome of trials. Both prosecution and defence teams regularly appeal to the emotions of jurors. In the United States, there is now an industry of psychological consultants who use personality tests to assist lawyers in rejecting jurors who may be unfavourably disposed to their position.

Another disturbing trend is the tendency for the legal system to become a redistributive mechanism (see also the discussion in Chapter 7). Juries appear increasingly likely to award damages on the basis of which party is more in need and which has the 'deeper pockets'. Juries are, however, instructed by judges about the appropriate reasons for damages awarded. Formerly, the allocation of property rights most likely to promote efficient resource allocation was the decisive factor.

More generally, the courts are often viewed as a type of insurance mechanism through which injured parties can receive payments to offset their losses. The court system is, however, an inefficient way to provide insurance. The costs of litigation are extremely high relative to the administrative costs of running a conventional insurance company. More importantly, altering the allocation of rights distorts incentives and can impose substantial costs on others in the future. The legal assignment of rights and responsibilities ought to be governed by notions of efficiency and justice, not attempts to achieve greater equality.

Some of these developments have arisen from within the legal system. Courts have been moving away from their traditional role of interpreting laws. Under pressure from political majorities to relax the constraints on what governments can do, courts have taken increasingly radical, or non-literal, interpretations of laws and constitutional constraints. Many politicians have fostered this tendency by encouraging courts to deal with controversial issues such as abortion rights, or the enforcement of norms of behaviour.

Legislatures also directly influence the legal system by introducing new statutes. A rational individual behind the veil of ignorance would want the legislature to overrule legal precedents using specific statutes:

> The fact that all law arising out of the endeavour to articulate rules of conduct will of necessity possess some desirable properties not necessarily possessed by the commands of a legislator does not mean that in other respects such law may not develop in very undesirable directions, and that when this happens correction by deliberate legislation may not be the only practicable way out. For a variety of reasons the spontaneous process of growth may lead into an impasse from which it cannot extricate itself by its own forces or which it will at least not correct quickly enough … There are several other reasons for this. One is that the process of judicial development of law is of necessity gradual and may prove too slow to bring about the desirable rapid adaptation of the law to wholly new circumstances. Perhaps the most important, however, is that it is not only difficult but also undesirable for judicial decisions to reverse a development, which has already taken place and is then seen to have undesirable consequences or to be

> downright wrong. The judge is not performing his function if he disappoints reasonable expectations created by earlier decisions. Although the judge can develop the law by deciding issues which are genuinely doubtful, he cannot really alter it, or can do so at most only very gradually where a rule has become firmly established; although he may clearly recognize that another rule would be better, or more just, it would evidently be unjust to apply it to transactions which had taken place when a different rule was regarded as valid. In such situations it is desirable that the new rule should become known before it is enforced; and this can be effected only by promulgating a new rule which is to be applied only in the future. Where a real change in the law is required, the new law can properly fulfil the proper function of all law, namely that of guiding expectations, only if it becomes known before it is applied. (Hayek, 1973: 88–89)

Legislative intervention can, however, also make the legal system less rather than more efficient. The political system will respond to many motives other than a general desire to increase the efficiency of resource use. Thus, legislation altering legal precedents may aim at benefiting vested interests just as taxation, spending or regulatory measures do.

Statutory laws reflect the balance of political power at the time they are enacted. The majority at any given time has an incentive to enact laws that favour its own vested interest at the expense of the long-term public interest. The majority is constrained only by constitutional laws, conventions or the expectation that whatever it does as the majority today might be done to it as the minority at a later date.

The expanding role of statutory law is, however, only part of the political influence on the legal system. Internationalisation of the legal system is another disturbing tendency. As governments sign various international treaties and conventions, domestic laws are increasingly being influenced by international bodies with little or no accountability to the people affected by those laws. It is ironic that public opinion in democratic nations is influenced by international organisations that are dominated by representatives of non-democratic governments. One suspects that some political activists are drawn to international institutions because these institutions are not accountable to anyone. When decision-makers do not have to answer to anyone, they are free to exercise the coercive power of government according to their own whims.

Liberty as an end in itself

Burke observed that the convention in Western democracies has been to regard our basic rights as fundamental and not as means to other ends:

> You will observe, that from the Magna Charta to the Declaration of Right, it has been the uniform policy of our constitution to claim and assert our liberties, as an entailed inheritance derived to us from our forefathers, and to be transmitted to our posterity; as an estate specially belonging to the people of this kingdom without any reference whatever to any other more general or prior right. (Burke, 1986: 119)

By contrast, the discussion thus far has focused on which institutions, or rules of social interaction, might be chosen by a rational individual behind the veil of ignorance who wished to further the goal of efficiency, moderated by a concern for avoiding poverty. A classical liberal would reject the assumption implicit in this view that people are, or would prefer to live in a society in which humans behave as, 'utility machines'.

Green (1996) provides a modern defence of the ideal of liberty that is based on the ideas expounded by classical liberal scholars. He distinguishes two modes of association:

> A 'corporate association' is composed of persons united in the pursuit of a common interest or objective … In a nation of 'civil associates' people are united not because they share a concrete goal, or are engaged together in a substantive task, but because they acknowledge the authority of the rules under which they live … In both types of association people are subject to rules of conduct, but in a corporate association the rules are instrumental to the pursuit of the common aim. In the pure form of civil association, the laws are moral stipulations, not instrumental commands. (Green 1996: 5)

Green notes that the idea of civil association rests on a particular view of human nature:

> Individuals are not seen as role players under leadership but as intelligent agents understanding or not understanding their predicament, guided by beliefs, choices, sentiments or habits. The freedom this entails is seen as a pleasure, not a burden. It is valued for its own sake, not simply because it may lead to wanted outcomes. Indeed, it is seen as central to the dignity of men and women. The individual is autonomous, not isolated … Individuals so conceived are the carriers of a moral compass: in short, each is a person with a character rather than an individual with a role. (Green, 1996: 6)

From the utilitarian perspective, a Kantian-type belief in the importance of individual liberty may also operate to limit the extent to which individuals can be exploited by a majority. Green (1996: 21) distinguishes what he calls an "habitual morality" from an "intellectual moral system". Whereas "an intellectual moral system erects rules or ideals to guide conduct … a man may [also] act morally out of habit, and his conduct may be recognisably good or bad, despite his lack of reflection". He goes on to assert that the habitual moral system is much more resilient against attack:

> Above all, such a moral system is the possession of the people, that is, of civil society. It stands or falls by the efforts of each person going about his or her daily life, upholding or not upholding the virtues it embodies. And this is its relevance for liberty. Because such a moral system rests on daily face-to-face practice, it is less prone to manipulation by the authorities. (Green, 1996: 21–22)

Hayek similarly emphasises that formal constitutional constraints are not the main forces limiting the exercise of coercive power. Rather, moral values that citizens inherit from previous societies keep the forces of tyranny, including the tyranny of a majority, at bay:

> [A] condition of liberty in which all are allowed to use their knowledge for their purposes, restrained only by rules of just conduct of universal application … is likely to be achieved and maintained only if all authority, including that of the majority of the people, is limited in the exercise of coercive power by general principles to which the community has committed itself. Individual freedom, wherever it has existed, has been largely the product of a prevailing respect for such principles which, however, have never been fully articulated in constitutional documents. Freedom has been preserved for prolonged periods because such principles, vaguely and dimly perceived, have governed public opinion. The institutions by which the countries of the Western world have attempted to protect individual freedom against progressive encroachment by government have always proved inadequate when transferred to countries where such traditions did not prevail. (Hayek, 1973: 55)

People need confidence that their basic rights will be guaranteed regardless of the individuals elected to govern. Epstein observes that the United States Constitution as originally proposed:

> … does not speak a word of individual substantive protections. Within the original framework the rich array of procedural and jurisdictional protections was expected to serve some substantive end. And that end was, of course, the protection of private property, of "lives, liberties and estates" that Locke considered the purpose of government. The procedural safeguards worked to guarantee by indirection that the government would not pass laws that encroached upon the property rights that government was designed to protect. "The Constitution," Hamilton wrote, "is itself, in every rational sense, and to every useful purpose, a Bill of Rights." (Epstein, 1985: 17)

But what if the confidence in the effectiveness of the array of "procedural and jurisdictional protections" is misplaced?

> With this background it is now possible to see how the Bill of Rights fits into the original constitutional scheme. The Bill identifies the ends of government, the

> rights that the system of limited jurisdiction, indirect voting, and separation of powers is designed to protect. (Epstein, 1985: 18)

A formal Bill of Rights, particularly one that is *not* chosen (or interpreted judicially) from behind a veil of ignorance, may in practice fail to protect, or may even erode, fundamental rights and freedoms. For example, Epstein (1996) notes three fundamental difficulties with New Zealand's Human Rights Act 1993:

> First … nobody claims that the entire corpus of New Zealand law dealing with rights enjoyed by individuals is contained in the Act. People possess many other rights, such as rights of conscience, rights of religion, rights of free speech, various rights of association, and even rights over one's own person and property …
>
> [S]econd … only certain individuals, occupying certain roles, can claim protection of the statute, while other individuals, occupying other roles, are unambiguously subjected by law to certain correlative duties … A meaningful human rights statute will surely protect all individuals equally, and equal protection of the laws … can only be afforded through some system of formal equality …
>
> [T]hird … [c]lassical jurisprudence always insisted that human rights came with human duties, which were the correlative to the rights created. Rights were never free goods; we always held a right against somebody else. To the extent that one person's liberty of action is expanded, the liberty of action of other individuals is necessarily limited. In designing a system of rights, the advantages that we conceive of and create for certain individuals should more than offset the disadvantages thereby imposed upon others. Any comprehensive system of rights should thus be couched within a framework of correlative duties. (Epstein, 1996: 4–6)

Epstein maintains that these defects imply that the Human Rights Act 1993 is contrary to the public interest because it is likely to produce inefficient outcomes. The Human Rights Act 1993 results in a systematic tendency to underestimate the costs of the explicit interventions, or indirect changes in behaviour, that it produces. Furthermore, by failing to uphold rights that are fundamental to prosperity, it may indirectly devalue those rights and thus endanger prosperity in the long run.

As a practical matter, however, one wonders whether appeals to efficiency will prove as effective against encroachments on fundamental rights and freedoms as would a more direct appeal to the inherent value of those rights. Majority rule, or even formal constitutional provisions chosen by some super-majority process thought to produce longer-lived legal constraints, cannot be relied upon to protect fundamental liberties. Liberty ultimately may only be sustained by a widespread belief in its value as an end in itself, and continual vigilance by its defenders to protect it against the desires of the governing majority of the day:

> It is the common fate of the indolent to see their rights become a prey to the active. The condition upon which God hath given liberty to man is eternal vigilance; which condition if he break, servitude is at once the consequence of his crime and the punishment of his guilt. (John Philpot Curran: Speech upon the Right of Election, July 10, 1790)

Horizontal equity as an end in itself

Another way to control possible abuses of majority rule is to treat horizontal equity as an end in itself. Recall from Chapter 2 that a law, or set of institutions, is horizontally equitable if it treats like individuals in the same way. Epstein (1993: 93–98) argues that horizontal equity may be justified on utilitarian grounds, as a way of avoiding the loss of resources that might otherwise be expended in attempting to obtain special treatment:

> … [when] the surplus generated … [by joint action] … may be allocated through political action in any way whatsoever … interested groups will strive to obtain the largest share of the surplus for themselves. These efforts cost time, money, and resources to mount. None of these efforts to reallocate the gain will increase social wealth; they will only decrease its overall size. At the limit it is possible that the full amount of the surplus could be dissipated by the efforts to direct it to one group or another, although in many cases the level of dissipation may be less, especially if a dominant faction is able to control political outcomes at low cost.

One could thus argue that each proposed case of favourable discriminatory treatment is likely to encourage destructive competition among groups seeking to become the next favoured party. Principled opposition to discrimination among people similarly situated may, however, be more effective at limiting the "tyranny of the majority". As we argued in Chapter 2, the opposition to discriminatory treatment appears in practice to be based on more than a concern for the resources wasted through lobbying activity.

Competition among governments

Competition among governments may assist in constraining some abuses of majority rule. When there are multiple local jurisdictions, people can vote with their feet. If people do not like the taxing and spending policies of one government, they can often move to another jurisdiction. The threat of losing population, and the associated tax base, then constrains government from pursuing the interests of its current majority supporters.

Multiple jurisdictions also permit experimentation with new policies. A radical policy can be tried in a context in which the damage from any unforeseen negative effects can be limited. Fewer people will be affected by the mistake, while those who are worst affected can move to another jurisdiction. If the new policy is deemed a success, it could later be adopted in other jurisdictions. Furthermore, when the costs or benefits of a policy extend beyond jurisdictional boundaries, local politicians will have little incentive to consider the consequences for people living in neighbouring jurisdictions. Unless adjoining jurisdictions can agree on policies or fiscal transfers, the outcome is unlikely to be efficient.

Dividing functions among federal, state and local governments may also constrain the exercise of political power by a majority since different majorities may prevail at different levels. Administrative costs, however, are likely to be increased through duplication of bureaucracies.

Furthermore, there are many examples of state and local governments implementing laws that benefit their own local vested interests at the expense of the wider community interest. For example, prior to Federation in both Australia and the United States, various states imposed tariffs on interstate trade. The fiasco resulting from railways being built with different gauges in the states of Australia is another example. State governments also use subsidies to compete against each other in attracting industries, particularly industries that are labour intensive. Many of the incentives offered to such firms will represent income transfers and a net efficiency loss to the community as a whole. For example, industries are likely to end up locating in areas that do not minimise production costs and inefficient investment projects may be undertaken while more efficient but less politically attractive projects fail to materialise.

One purported advantage of multiple levels of government is that local governments are closer to the people. We noted above that information costs are a barrier to effective government policy. It is conceivable that people are likely to have much better information on the effects of policies implemented at the local level. On the other hand, with less at stake, voters may pay less attention to local politics. Certainly, politics at the local level is commonly perceived to be susceptible to corruption.

In the New Zealand context, the foremost example of jurisdictional competition is that between New Zealand and Australia due to free migration of people between these countries. Such potential migration constrains the policies that any government in the region can implement.

Conservatism and social change

A rational person behind the veil of ignorance might insist on a degree of conservatism in social and political affairs. Stability of rules and institutions gives people some confidence in the decisions that will be made by others. Individuals can make their own decisions, and plan their own future actions, with greater confidence if rules are stable. Furthermore, it is difficult to know all the consequences of a policy change. Rules or institutions that have worked well in the past should be adapted before more radical changes are attempted. Social scientists need, however, to remain open to the possibility that the initial expectations of the effects resulting from a particular change may be mistaken.

CONCLUSION

As Winston Churchill said in a speech to the House of Commons in November, 1947, "Democracy is the worst form of government except all those other forms that have been tried from time to time". Democracy accompanied by the (largely) independent rule of law is far from a perfect social decision-making mechanism. In particular, there is no reason to suppose the outcome of democratic government will be all and only those social rules of interaction that might be judged ethical by a process of rational discussion and argument. Ethical discussion, and social science generally, nevertheless are relevant to politics. Common ethical standards, and other shared beliefs, influence both political and legal decisions.

Compassion for the poor, or a feeling of sympathy for those in distress, is integral to our ethical system. Private charity is one means of accommodating such concerns without compromising efficiency or horizontal equity. Unfortunately, private charity alone may not suffice. There may still be grounds for some degree of compulsion in offering assistance to people living in poverty.

In a modern market economy, a minimal level of money income is needed to obtain an acceptable standard of living. Thus, we argue in Chapter 5 that a policy of redistributing to people in dire circumstances does not require that we support the goal of greater equality. Like private charity, a policy of redistribution also follows from a desire to act compassionately.

In practice, there does seem to be widespread support for the notion that the genuine poor should be helped. More accurately, there is support for good intentions with regard to helping the genuine poor. The concern may not extend to ensuring that anything is actually achieved for those in need.

A related question is whether government ought to be allowed to redistribute income to anyone but the poor. It is not surprising that the combination of majority rule and a redistributive role for government has led to a so-called 'churning state'. Most of the beneficiaries (apart from the suppliers) of many government programmes, such as public education or publicly subsidised health care, are from the middle class, the same people who pay most of the taxes. There is a churning of funds from the middle class back to the middle class with considerable waste and inefficiency, and substantial creaming of rents for suppliers, along the way. Most of us would be better off if we constrained ourselves from being able to vote for such policies in the first place.

A belief in equity is often taken to mean a belief in the value of more equal outcomes as an end in itself. The only defensible measure of outcomes is welfare, which is impossible to measure. In practice, annual money income is often used as a proxy for welfare even though, as we shall argue in Chapter 5, there are many reasons to doubt that the correlation between annual money income and welfare is high.

APPENDIX: RISK AVERSION

This appendix outlines how we can use expected utility theory to model the trade-off between efficiency and the risk of a low standard of living. The mathematical apparatus we outline in this appendix also underlies some of our later discussions of different measures of inequality.

Expected utility theory is the standard model of choice in uncertain situations. The applicability of expected utility theory has been tested in many economic environments, and in psychological experiments in which the theory has performed less well. Relatively minor adjustments to the theory can, however, account for many of its inconsistencies with the empirical evidence. In particular, expected utility can closely approximate more general preferences. There are other situations in which more radical modifications to expected utility are needed to account for the evidence. The strengths and weaknesses of expected utility theory are, however, not important for our current purposes.

The basic idea behind expected utility theory can be explained by imagining that an individual is attempting to determine the value of a set of mutually exclusive random outcomes. Let the possible outcomes, for example different lifetime monetary incomes or wealth levels, be a finite number of values $y_1, y_2, \ldots, y_n$. We shall argue in Chapter 5 that we are not likely to be able to find a single index of living standards, such as wealth or expected lifetime monetary income, but that issue is not important for our current discussion.

Assume that the value or 'utility' of outcome y_i can be represented by a function $U(y_i)$ and suppose the individual prefers more to less, so that U increases as y increases. Equivalently, the marginal utility of additional wealth at any current wealth level:

$$U'(y) \equiv \frac{dU}{dy} > 0 \qquad \text{(EQ 3.1)}$$

is positive. Assume also that the individual believes that outcome y_i can occur with probability p_i. The *expected utility* of the set of possible outcomes is then the utility of each outcome multiplied by the probability of that outcome occurring, with the result then summed over all outcomes in the set:

$$V = \sum_{i=1}^{n} p_i U(y_i)\,. \qquad \text{(EQ 3.2)}$$

(Equation (3.2) can easily be extended to a situation in which the possible outcomes are a continuous variable.)

A key advantage of expected utility theory is that properties of the utility function U capture attitudes to risk and reward. On the other hand, the attractive, and unattractive, features of a random choice are represented by properties of the probability distribution, in this case the values of y_i and p_i.

Expected utility and risk aversion

There is considerable evidence that most people are risk averse. More risky financial assets need to pay a higher expected return in order to compensate investors for bearing the additional risks. There is a substantial market for insurance against many adverse events including fire, theft, illness or death. Workers need to be paid a higher wage on average before they willingly consent to work in a more dangerous job or situation.

Many people gamble, and since these same people buy insurance and display risk aversion in their financial dealings, gambling *per se* may not reflect risk-loving behaviour. Rather, gambling may provide other benefits such as entertainment, the social value of participating in a major sporting or national event (such as the Melbourne Cup in Australia) or the psychological benefits of dreaming about what one might do with additional money. We discuss another feature of gambling later in this appendix, since the characteristics of gambles that people find attractive may give us important information about attitudes to risk.

Risk aversion corresponds to a property of the U function known as concavity, or decreasing marginal utility of wealth. An individual is *risk averse* if the loss in utility associated with a decline in wealth is greater in magnitude than the gain in utility associated with an increase in wealth of the same size. Losses are therefore avoided more than similar-sized gains are sought.

Since the marginal utility is the first derivative of U with respect to y, the change in marginal utility depends on the second derivative of U with respect to y. For the effect of a change in wealth to be greater at a lower wealth level, as is the case for a risk averse person, the second derivative of U with respect to y has to be negative:

$$U''(y) \equiv \frac{d^2U}{dy^2} < 0 \,. \tag{EQ 3.3}$$

Means and variances

Recall that we are imagining an individual behind the veil of ignorance who is contemplating a finite set $y_1, y_2, ..., y_n$ of potential lifetime monetary incomes or wealth. This person further believes that outcome y_i can occur with probability p_i.

The *subjective mean*, or *expected*, wealth will be

$$\mu = \sum_{i=1}^{n} p_i y_i . \qquad \text{(EQ 3.4)}$$

Increasing expected wealth as defined in equation (3.4) then corresponds to raising efficiency as we defined that concept in Chapter 2. If expected wealth increases, decreases in wealth in any subset of possible outcomes must be more than compensated for in expected value terms by increases in other situations. The winners who gain in wealth could compensate the losers in an expected value, or *ex ante,* sense and still gain wealth themselves. Maximising expected value in this context corresponds to increasing average wealth while ignoring its distribution.

In other contexts, wealth maximisation does not equate to maximising efficiency. In particular, wealth as commonly measured takes into account only the expected present value of income flows that accrue to assets that can be bought or sold. Maximising wealth then would not correspond to maximising the expected present value of overall expected income flows. Furthermore, the above discussion assumes implicitly that lifetime monetary incomes or wealth completely determine individual utility levels. This is not accurate, since people care about more than the market goods and services they consume. These distinctions are not important for our present purposes, however, so we shall not discuss them further in this appendix.

Risk-averse people will care about more than the average wealth level. The *subjective variance* of the distribution of incomes will be

$$\sigma^2 = \sum_{i=1}^{n} p_i (y_i - \mu)^2 . \qquad \text{(EQ 3.5)}$$

The *standard deviation* σ is the square root of the variance. An income distribution A is said to be more risky than another distribution B with an equivalent expected value if any risk-averse investor would prefer distribution B to distribution A. Rothschild and Stiglitz (1970, 1971) have shown that this is equivalent to asserting that distribution B can be spread out (in a way that preserves mean values) to yield the distribution A. When the variances of the distributions A and B both exist, the variance of A will exceed the variance of B.

Attitudes toward risk

Individuals considering risky income streams tend to focus in the first instance on the average values of the possible outcomes. Risk aversion becomes important only as a second order effect, even for individuals whose degree of risk aversion is very large.

This result can be illustrated by imagining an individual contemplating taking on a risk when initially in a non-risky situation. Specifically, suppose a risk-averse individual behind the veil of ignorance is currently in a situation of complete equality. The possible income levels $y_1, y_2, \ldots, y_n$ discussed above all have the same value y. The individual is now offered a chance to increase mean income above y by accepting a source of income variation that is risky. Let the new source of income have a mean $\mu > y$ but a variance that is positive (so it has income levels, say $x_1, x_2, \ldots, x_n$, that differ in the n possible situations the individual was originally contemplating). Assume that in order to take on an amount z of the new risk the individual has to give up an amount z of the original income y. The possible total incomes in each situation then become:

$$(1-z)y + zx_1, (1-z)y + zx_2, \ldots, (1-z)y + zx_n \qquad \text{(EQ 3.6)}$$

and if x_i is close to zero for some value i, taking on the risk could reduce the minimum income level the individual may experience in situation i. The expected utility as a function of z would be:

$$V(z) = \sum_{i=1}^{n} p_i U[(1-z)y + zx_i]\,. \qquad \text{(EQ 3.7)}$$

The question in which we are interested can now be stated as follows. When would the individual choose a value for z that is strictly positive? So long as U is differentiable, we can evaluate the derivative of V with respect to z at the value $z = 0$:

$$V'(0) = \sum_{i=1}^{n} p_i U'(y)(x_i - y) = U'(y)\sum_{i=1}^{n} p_i(x_i - y) = U'(y)(\mu - y)\,. \qquad \text{(EQ 3.8)}$$

A decision to accept a small amount of the risky income stream depends only on the marginal utility of income at the initial level y and the mean of the risky income stream relative to y. Since $V'(0) > 0$, the individual could become better off in expected utility terms by choosing a value for z that is greater than zero. An individual can have an arbitrarily large, albeit finite, degree of risk aversion and yet still accept some of the risky income stream, even though the minimum possible income could decline as a result.

The intuitive explanation for this result is that expected utility depends linearly on average income but quadratically on income variability. Specifically, we can approximate $V(z)$ by:

$$V(z) \cong V(0) + V'(0)z + \frac{1}{2}V''(0)z^2 \qquad \text{(EQ 3.9)}$$

and, although $V''(0) < 0$ for a risk-averse person, for small values of z the value of z^2 is much smaller than z (for example, when $z = 0.2$, $z^2 = 0.04$). Thus, no matter how great the degree of risk aversion (so long as it is finite), some positive value of z would maximise expected utility. An individual who is risk averse, but not infinitely risk averse, will nevertheless find higher expected value more attractive than additional risk up to some point.

Maximin and risk aversion

Rawls assumes that an individual behind the veil of ignorance would maximise the value of the worst possible outcome. Such a person would therefore forgo the risky opportunity discussed in the previous section if x_i for any one outcome $i = 1, 2, \ldots, n$ is less than y. A person who maximises the worst outcome is said to have a maximin objective. Since the results above do not hold for such a person the corresponding utility of income function $U(y)$ cannot be differentiable. Also, from (3.9), we can see that risk can be of overwhelming importance if only the decision-maker effectively is infinitely risk averse. In such a case, the infinite coefficient on the quadratic term in (3.9) allows that term to dominate.

The equivalence between a maximin objective and infinite risk aversion might be seen most easily when there are just two possible outcomes y_1 and y_2. If the probability of attaining y_1 is p and the probability of attaining y_2 is $1–p$, the expected utility function (3.2) becomes:

$$V = pU(y_1) + (1-p)U(y_2)\,. \qquad \text{(EQ 3.10)}$$

Combinations of y_1 and y_2 that yield equal levels of expected utility, or 'expected utility indifference curves', can then be drawn as in Figure 3.2. The slope of one of these curves can be found, by differentiating (3.10) for a constant level of V, to be:

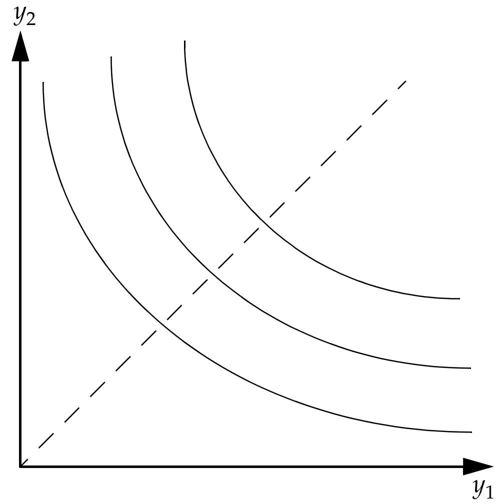

FIGURE 3.2: Expected utility indifference curves for a risk-averse individual

$$\frac{dy_2}{dy_1} = -\frac{p}{1-p}\frac{U'(y_1)}{U'(y_2)}. \qquad \text{(EQ 3.11)}$$

The change in this slope as y_1 increases follows from differentiating (3.11) with respect to y_1 and using (3.11) to substitute for dy_2/dy_1:

$$\frac{d^2 y_2}{dy_1^2} = -\frac{p}{1-p}\left\{\frac{U''(y_1)}{U'(y_2)} + \frac{p}{1-p}\left[\frac{U'(y_1)}{U'(y_2)}\right]^2 \frac{U''(y_2)}{U'(y_2)}\right\}. \qquad \text{(EQ 3.12)}$$

Risk aversion, as in (3.3), then implies that indifference curves get flatter (the sign of the second derivative in (3.12) is positive) as y_1 increases. The higher the degree of risk aversion, the faster will the indifference curves flatten out and the more the indifference curves will be 'bowed in' toward the origin. The limit of this process is infinite risk aversion when the indifference curves become right angles as in Figure 3.3. In the case illustrated in Figure 3.3, only the minimum value of y_1 or y_2 matters to the level of expected utility.

Maximin is a special limiting case where risk aversion becomes infinite. Furthermore, it is a special case that is of no empirical relevance. Degrees of risk aversion can be estimated from many types of economic data including risk premiums paid on assets, the levels of insurance people purchase, the amount they wager in gambles and the wage premiums demanded for more

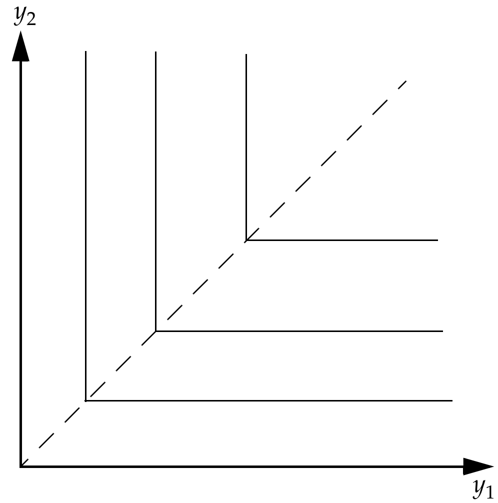

FIGURE 3.3: Indifference curves for a maximin individual

risky occupations. In all cases, infinite risk aversion is easily rejected by the evidence.

Adding to existing risks

The above discussion related to accepting a risk starting from a situation of complete certainty. A new risk could also be assumed by an individual who is already facing an uncertain set of possible future incomes. Suppose this new risk has a higher mean, but also a higher variance than the existing risky income distribution. In this case, too, the attraction of a higher mean on a new risk is likely to dominate the undesirable effect of a higher variance for all but the most risk-averse individuals.

To begin with, unless the new risk and the existing distribution are perfectly correlated, the standard deviation of a weighted average of the two income flows will be less than the weighted average of the two standard deviations. If the two income flows are less than perfectly correlated, when one income flow is substantially below (or above) its mean the other may be less so. If the correlation is low enough, even an infinitely risk-averse person would choose to take on some of the new risk to exploit the advantages of diversification. In addition, the higher mean of the new risk will also make it attractive to all individuals except those who are infinitely risk averse.

The situation can be illustrated as in Figure 3.4. In this illustration, we have assumed that the income distributions are all normally distributed so that we can use the mean and variance to characterise them. If the mean and standard deviation of distribution A are μ_A and σ_A while the mean and standard deviation of B are $\mu_B > \mu_A$ and $\sigma_B > \sigma_A$ then the mean of a weighted combination of A and B is:

$$\mu = (1-z)\mu_A + z\mu_B . \qquad \text{(EQ 3.13)}$$

Suppose we again use $y_1, y_2, \ldots, y_n$ for the possible income levels under distribution A and $x_1, x_2, \ldots, x_n$ for the corresponding income levels under distribution B. Analogously to (3.5), the covariance between distributions A and B is defined as:

$$cov(A, B) = \sum_{i=1}^{n} p_i(y_i - \mu_A)(x_i - \mu_B) \qquad \text{(EQ 3.14)}$$

and the correlation between them is:

$$\rho = \frac{cov(A, B)}{\sigma_A \sigma_B} . \qquad \text{(EQ 3.15)}$$

The standard deviation of a weighted average of A and B can then be written:

$$\sigma = \sqrt{[(1-z)^2\sigma_A^2 + z^2\sigma_B^2 + 2z(1-z)\rho\sigma_A\sigma_B]} . \qquad \text{(EQ 3.16)}$$

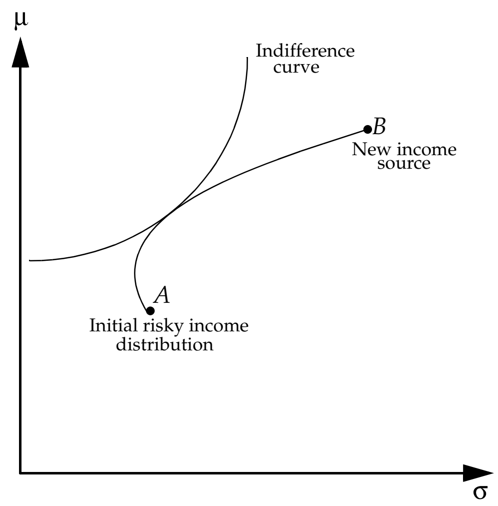

FIGURE 3.4: A diversifying source of income variation with higher mean and variance

The curved set of possible income distributions in Figure 3.4 is the graph of μ from (3.13) against σ from (3.16) as z varies from zero to one. If the two income distributions were perfectly correlated, the distributions that could be obtained by trading some of the initial income source for some of the new one would be represented by the straight line joining points A and B. If the two sources of random income are less than perfectly correlated, adding some of the new source will generally allow risk to be reduced while mean income is simultaneously raised. The frontier will be a parabola and will bend backward at A if:

$$\rho < \frac{\sigma_A}{\sigma_B}. \qquad \text{(EQ 3.17)}$$

Since the right side of (3.17) is positive, the frontier can bend back even if the two income sources are positively correlated. Thus, diversification benefits are obtainable even from distributions that are positively correlated so long as the correlation is not too close to the maximum upper bound of one.

A risk-averse expected utility maximiser will like an increase in mean income μ, dislike an increase in standard deviation σ and have indifference curves running southwest to northeast and shaped like the one drawn in Figure 3.4. An infinitely risk-averse person would care only about σ and would have indifference curves that are vertical straight lines.

The risk-averse person whose preferences are represented in Figure 3.4 would choose to trade some of the safer distribution A for the more risky one B. Even an infinitely risk-averse person would accept some of the new income source for the correlation illustrated in Figure 3.4. More generally, even though the frontier may not bend backwards at point A it is likely to be quite steep. A person would have to be very risk averse not to trade any of the original income source for some of the new one.

Attitudes to skewness

The *skewness* of the distribution of incomes $y_1, y_2, \ldots, y_n$ given above is conventionally defined as the third moment about the mean normalised by the standard deviation:

$$s^3 = \sum_{i=1}^{n} p_i \left(\frac{y_i - \mu}{\sigma}\right)^3. \qquad \text{(EQ 3.18)}$$

Since the cube of a negative number is negative, values of y below the mean contribute negatively to the skewness, while values above the mean contribute positively. If the distribution is symmetric about the mean (for example, it is normal), the skewness will be zero as the positive and negative terms cancel. A distribution will have positive skewness when large values above the mean are more likely to occur than are small values equally far below the mean. Conversely, a distribution biased towards low values will have a negative skewness.

Evidence derived from insurance, investment and gambling behaviour suggests that skewness, when it is non-zero, may be almost as important as variance with regard to attitudes toward risk. For example, when buying insurance people seem to be most concerned with catastrophes. They are willing to accept policies with large deductibles that do not cover smaller adverse events. Such preferences for insurance policies may reflect a dislike of negative skewness (although an alternative explanation, based on the administrative costs and incentive effects of insurance, is discussed in Chapter 4).

Investors in financial assets also seek positively skewed, and avoid negatively skewed, returns. Options contracts are more desirable than futures contracts as hedging instruments, precisely because options allow the investor to avoid large negative outcomes while retaining the benefits of large gains. In contrast, futures contracts eliminate the potential for gains as well as losses. An investor concerned only about variability of returns would prefer the futures contract.

Finally, gambling behaviour also suggests that people seek positively skewed outcomes. Lotteries and other gambles typically pay large amounts of prize money with a small probability. If the large prizes were split into many smaller prizes, each of the smaller prizes could be won with a higher probability. The resulting gamble would be less risky, but also less skewed. The fact that people prefer the more risky gamble suggests a significant demand for positive skewness in outcomes.

Golec and Tamarkin (1998) provide more systematic evidence on the demand for positively skewed outcomes. They examine betting on horse races and find that bets with a low probability of paying off and a high variance of payoff amounts (so-called long shots) provide low mean returns. Bets on favourites, that have a relatively high probability of paying off and a low variance of payoff amounts, provide relatively high mean returns. Golec and Tamarkin show that bets on long shots have extremely large positive skewness. Evidence is thus consistent with the idea that gamblers are risk averse but have a strong desire for positively skewed outcomes. The high demand for bets with a large positive skewness reduces the average return on those bets relative to the bets on favourites.

A minimally acceptable standard of living

A risk-averse person behind the veil of ignorance would not have a strong preference for income equality. A person who was anything less than infinitely risk averse would exchange a guaranteed fixed income for a risky income distribution that has a higher mean value. A person who was anything less than infinitely risk averse also would be likely to trade some of a risky income for another risky income that has a higher mean, particularly when the two income flows are less than perfectly correlated. The new income flow then provides diversification benefits. Even a risk-averse person is likely to find a positively skewed income distribution attractive, despite that distribution also having a high variance.

While equality may be of no interest to a risk-averse person behind the veil of ignorance, such a person would be concerned to avoid poverty. Income distributions that include a significant chance of poverty may be negatively skewed. A desire for positive skewness corresponds to a dislike of negative skewness. Eliminating low income levels is one way of increasing the positive skewness of an income distribution.

In addition, a dislike for poverty may be a different type of concern from attitudes to risk or skewness in income levels that are guaranteed to remain above the poverty level. Specifically, people may require lifetime monetary income, or, in a situation in which people cannot borrow against future income, a monetary income over a shorter time horizon, to remain above a minimally acceptable level. Specifically, suppose there is some level of income y_{min} so that:

$$U(y) \to -\infty \text{ as } y \to y_{min} \qquad \text{(EQ 3.19)}$$

but $U(y)$ is finite for all income levels in excess of y_{min}. Suppose further that someone is facing the chance of earning y_{min} with a non-zero probability. A small increase of income in that situation would bring an infinite gain in utility. The individual thus would be willing to sacrifice a large amount of income in other situations if it enabled this individual to avoid having $y = y_{min}$ in that one particular situation.

Aversion to poverty is, however, different from aversion to risk. As long as income is guaranteed to remain above y_{min} in all possible situations, an individual may be willing to take on a substantial amount of risk in order to raise expected income or positive skewness in return.

4

Equality and Efficiency

INCREASING EFFICIENCY AND EQUALITY

Efficiency and equality are often thought to be competing goals. In particular, one argument against reallocating resources to make someone better off is that the reallocation could exacerbate inequality. There are many policy changes, however, that could be expected both to increase efficiency and to reduce inequality. These involve eliminating many policies that reduce efficiency and economic growth by granting legally protected privileges of one sort or another to politically favoured groups:

> Though a by-product of freedom, greater equality is not an accident. A free society releases the energies and abilities of people to pursue their own objectives. It prevents some people from arbitrarily suppressing others. It does not prevent some people from achieving positions of privilege, but so long as freedom is maintained, it prevents those positions of privilege from being institutionalised; they are subject to continued attack by other able, ambitious people. (Friedman and Friedman, 1979: 148)

As our discussion of the marginalist theory of income distribution in Chapter 2 illustrated, competition tends to eliminate rents, where rents are defined as differences between the total value produced by an allocation of resources and the total value produced by the next best alternative allocation. High incomes attract entry by competitors. The best way to ensure continuing high incomes, therefore, is to erect barriers to entry. In turn, the most effective barriers to entry are legal prohibitions.

In a small open economy such as New Zealand, foreign producers provide the greatest competition to domestic suppliers. Eliminating barriers to trade promotes efficiency while simultaneously reducing a major source of rents.

Labour market restrictions also reduce efficiency and generate rents. These restrictions include occupational licensing, artificial credentials and legal privileges for trade unions. Some labour market policies may appear to help the poor, but in fact tend to have the opposite effect. For example, minimum wage laws seem to benefit less-skilled members of the work force. In practice, however, these laws make it harder for less-skilled or less-experienced workers to find a job. The law allows people to work for nothing, or for the minimum or a higher wage, but not for any wage between zero and the minimum.

As argued by Hartley (1992) and others, the adverse effect of minimum wages on employment can be moderated by adjustments that employers can make in other aspects of a job. The minimum wage specifies a payment per hour of time spent on the job. The employer is interested, however, in the amount of labour services actually provided. The relationship between hours of work and labour services can be altered in many ways. Assembly lines can be run faster, workers can be allowed fewer or shorter work breaks, workers can be supervised more closely, less time can be devoted to keeping the work place pleasant or clean, non-wage benefits can be reduced and so forth. The situation is analogous to the effect of rent controls. Just as rent controls discourage landlords from maintaining the quality of rental housing stock, so minimum wages discourage employers from offering high-quality jobs. The result is that the trade-off between the wage and non-wage characteristics of jobs is less efficient than unrestrained bargaining between employers and employees is likely to produce. The net effect of minimum wages on employment is, however, reduced to the extent that other, less undesirable, margins of adjustment are available. Nevertheless, the most reputable economic studies have shown there is a statistically significant, albeit small, negative effect of minimum wages on employment.

Labour market studies have shown that prior work experience is one of the most important factors in determining current wages. When people are prevented from obtaining entry-level jobs, their future market incomes may be permanently reduced. Unemployment benefits may lessen the immediate negative effects of minimum wages on the unemployed, but such benefits come at a cost of reduced efficiency and prosperity for the country as a whole. Unemployment benefits are also a poor substitute for work experience and higher future wages as far as benefit recipients are concerned.

Minimum wage laws may also prevent workers from obtaining additional skills. Workers can invest in training by working for an employer at a lower wage. The lower wage can be considered as payment for services provided, minus a training charge returned to the employer. Competition would not eliminate the training charge because employers face higher costs while employing untrained workers. Not only are there explicit training costs, such as paying for increased supervision, but employees are also likely to be less productive before they are trained.

On-the-job training will always be paid, in part, by the trainees because it prepares them for jobs outside of the firm providing the instruction. A trained worker can obtain a higher market wage because newly acquired skills can be used in another job. The worker's current employer therefore will have to raise the worker's pay in order to compete. Training costs thus can only be recovered by the employer through payment of a lower-than-market wage during the training period.

The point may be clarified by contrasting 'general' training with 'job specific' training. Training that produces skills that are specific to one employer does not raise the amount other employers are willing to pay for the worker's services. The trained worker's market wage rises by a lesser amount than does productivity in the current job. The employer's share of training costs therefore can be recovered in subsequent wages that are less than the worker's productivity.

More skilled or more experienced workers may be the main beneficiaries of a minimum wage. An increase in the price of unskilled and inexperienced labour makes alternative production technologies more attractive. The demand for skilled workers should increase as a result. By reducing the ability of unskilled and inexperienced workers to invest in skills and experience, minimum wages may also reduce the supply of skilled workers below what it would otherwise be. An increased demand, and reduced supply, of skilled labour should thus raise the wages of current skilled and experienced workers. Although touted as a means of helping the poor, minimum wage laws are a sectional policy aimed at benefiting the skilled, experienced and politically organised workers at the expense of the least-advantaged members of society.

Once we have removed barriers to competition in the traded goods industries and the labour markets, we could move next to the non-traded goods industries that have been monopolised by the government. These include many of the infrastructure industries critical to economic growth and prosperity, such as communications, energy and transport services. Governments made many of these potentially competitive industries into unnatural monopolies. In many cases, the monopoly producers were established as publicly owned firms in which competition in ownership, with its attendant disciplines on managers and employees, was eliminated. Only in recent years, with deregulation and privatisation, has this situation begun to change.

A related set of inefficiencies is associated with industries, such as education and health care, in which there are no monopolies in the usual sense but in which supply is provided by public enterprises. We discuss these sectors, particularly education, in more detail in Chapter 6. The resulting poor quality of public education services often adversely affects the indigent and children whose parents are less well educated.

Many other policies help particular interests at the expense of the general public interest, often to the detriment of the poor. Land use and zoning policies often benefit middle- and higher-income groups by preventing poorer people from taking advantage of the better amenities available in wealthier suburbs. Zoning laws often provide large returns for land speculators. Many wealthy individuals have benefited from buying land zoned for one type of use and then persuading a government entity to re-zone the land for an alternative, more valuable, use. It is significant, for example, that large land holders have been some of the strongest proponents of introducing zoning laws into Houston, Texas (perhaps the largest city in the developed world without zoning laws).

In many developing countries, price controls on foodstuffs aimed at urban voters effectively tax many of the much poorer small farmers. Conversely, in many developed economies, restrictions on agricultural production and trade often benefit very large agribusinesses at the expense of the urban poor.

There is no shortage of policy reforms that could increase both efficiency and equality. Many of these have been advocated in New Zealand by the New Zealand Business Roundtable and others. If policy-makers were truly concerned with efficiency and equity, such advice would have been received with open arms. We discussed in Chapter 3 how the incentives inherent in democratic politics may explain why this has not always been the case.

EFFICIENCY AND INCOME TRANSFERS

There is another sense in which conflict between the goals of efficiency and vertical equity are less serious than is commonly supposed. Economists have suggested a number of reasons why limited compulsory transfers of income by government might also be efficient. These arguments are based on the notion that various public goods problems, externalities or other so-called market failures may justify a certain amount of income redistribution. The amount of redistribution justified by these arguments is limited, however, and is directed at those most in need.

Public concern about poverty

Private transfers of income to the poor may provide external benefits to people other than the donor. Many individuals dislike seeing fellow citizens who are homeless, ill, poorly clothed and under-nourished. Any redistribution to the poor will provide benefits to people who are concerned about them. A person who does not help the poor may feel good when the poor are receiving aid despite that person's own reluctance to contribute. People 'free ride' on the income transfers of others.

Each person would voluntarily transfer income up to the point at which the marginal private benefits of an extra dollar transferred match the marginal costs in terms of the forgone private expenditure that could have been financed instead. The total community benefits of each dollar transferred are, however, the sum of all the benefits provided to each person who values such transfers. As a referee pointed out, we ought also to subtract the negative benefits suffered by the "social Darwinists in our midst who think that redistribution hurts the fabric of the race". Casual empiricism suggests, however, that there are more compassionate people than social Darwinists, while the strength of feeling of the former individuals exceeds the depth of concern of the latter. The net effect on other people of the last dollar of income privately transferred is therefore likely to be positive. An additional dollar of transfer income above the level chosen privately would then yield benefits above the marginal cost. Uncoordinated action by private individuals therefore might lead to less than the efficient amount of income redistribution. (See Roberts (1984) for a detailed argument.)

The argument that private redistribution would be inefficient is premised on there being private benefits associated with income transfers. The observation that people voluntarily give money to others is consistent with this claim. The characteristics of private charitable contributions suggest, however, that donors are not concerned about inequality as such. Rather, they care about poverty, or more particularly homelessness, inadequate clothing or nourishment, or inadequate access to health care, education or legal representation.

Income or expenditure insurance

As we observed in the appendix to Chapter 3, there is considerable evidence that most people are risk averse. If people are risk averse, they would be made better off if they could buy insurance against unforeseen fluctuations in their income or unavoidable expenditures. Since many of the random events that adversely affect one person's income or unavoidable expenditures are uncorrelated with the events affecting other people, it would in principle be feasible to provide insurance against such events.

Private insurance markets allow people to insure against some of the events that adversely affect their income or expenditure. One can purchase insurance that covers expenditures associated, for example, with illness or automobile accidents. Private annuity markets also provide income in retirement. More generally, private savings, and the ability to borrow against future earnings, are a form of self-insurance against adverse fluctuations in either income or expenditure.

Private insurance against some fluctuations in income or expenditure is, however, expensive if not impossible to obtain as a result of what economists call *moral hazard* and *adverse selection*. The idea of moral hazard first arose when actuaries observed that the probability of an adverse event occurring appeared to rise after insurance was made available. The basic idea is that indemnifying against losses reduces the incentive to avoid the adverse event. If it costs an individual any time or other resources to take care, insurance will inevitably tend to reduce the amount of care that is taken.

People can take many actions that affect their income or expenditure. Some of these actions are very difficult or expensive for anyone else to monitor or control. Income could be low because of bad luck, but it could also be low because a person has decided to take a little more leisure on the job. Expenditures may be unusually high because of unanticipated visits from friends or relatives, or a household can be secretly inviting friends and relatives to visit at the expense of the insurance company. People may remain unemployed for long periods because there are few suitable jobs available that match their skills, or because they are not very diligent in searching for potential employment. It may be difficult to distinguish these alternative causative influences.

Suppose income or expenditure insurance were provided to someone who has substantial control over the probability that the indemnified events will occur, or the magnitude of the restitution if the event does occur. If it is difficult or expensive to observe or measure the amount of care the insured person has taken, the insurance will be very expensive. The premiums would reflect the expectation of the insurance company that the insured individuals would respond to the incentives to increase the likelihood of claims or to inflate settlement amounts.

People who are less able to manipulate their affairs will find insurance an unattractive proposition. A company that offered an income or unusual expenditure insurance policy would suffer from adverse selection. The pool of willing customers would be adverse from the point of view of the likelihood of making large payments on the policy. The insurance company is likely to conclude that people willing to buy such insurance policies must have particularly unusual opportunities to manipulate their affairs.

Insurance companies might conclude that offering income or unusual expenditure insurance would be tantamount to volunteering to be exploited.

A profit-maximising insurance company would not want to sell a policy indemnifying unusual fluctuations in income or expenditure to anyone who would be happy to buy it.

In practice, income or expenditure insurance is only available for events that can be objectively observed and verified. For example, a virtuoso violinist can insure against the loss of a finger, while many people insure against extraordinary expenditures resulting from illnesses, accidents or fires.

Viewed from this perspective, redistribution can be seen as a type of generalised income insurance. Additional tax payments made when one's income is temporarily above average are analogous to an insurance premium. Income transfers received when one's income is temporarily below average are similar to an insurance indemnity payment.

A government offering such an insurance scheme is no less subject to moral hazard than a private insurance company would be. The existence of the tax and transfer mechanism reduces incentives to produce market income and increases incentives to place oneself into a category of those eligible to receive one of the transfers.

A government has one advantage over a private insurance company, however. Since the government can legally compel people to pay taxes when their incomes are positive (ignoring the possibilities of emigration, suicide or tax evasion), the government does not face the same extreme form of adverse selection. Essentially, it can force much of the working age population to participate in the scheme. The compulsion comes, however, at a high cost.

The efficiency cost of taxation

Taxation is inefficient because it drives a wedge between the buying and selling prices of a taxed good or service. The value of the good or service to the buyer then exceeds the cost for the seller by the amount of the tax. Production and consumption of the taxed good or service will be lower than it would have been without the tax. The losses incurred by the market participants will exceed the revenue raised by the government. This is illustrated in a simplified way in Figure 4.1 where E represents the market equilibrium in the absence of taxation, t is the tax rate, $p(1+t)$ is the tax-inclusive price paid by consumers of the good or service and Q is the output under taxation. The price p that the suppliers receive after the tax is imposed is lower than the equilibrium price in the absence of taxation.

The lighter shaded area bounded by p, B, E and F represents the cost of the tax to suppliers. The first component of this cost is the revenue they lose.

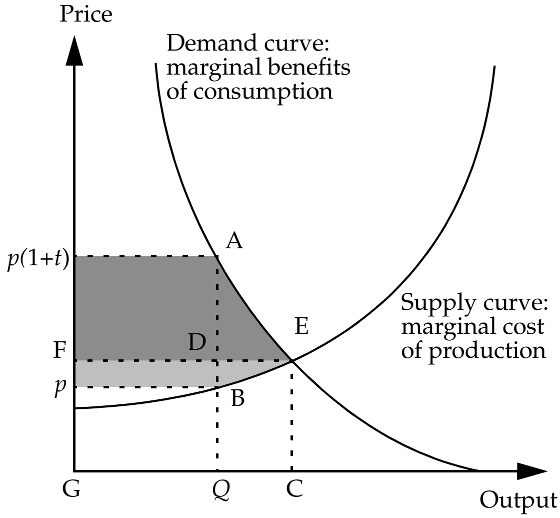

FIGURE 4.1: The efficiency costs of taxation

They lose pFDB on the Q units they continue to sell, together with all the revenue they formerly earned by selling more than Q units (the area bounded by D, E, C and Q). The reduction in output, however, also allows producers to reap some cost savings. These equal the area under the marginal cost, or supply, curve bounded by B, E, C and Q.

The darker shaded area bounded by $p(1+t)$, A, E and F represents the cost of the tax to consumers. Before the tax was imposed, consumers were enjoying benefits equal to the area under the marginal benefit curve above line FE. This area represents the total benefits from consumption level C minus the expenditure FECG. After the tax is imposed, the benefit for consumers net of expenditure becomes the area under the marginal benefit curve above line $p(1+t)$, A. The change in net benefits is thus the area bounded by $p(1+t)$, A, E and F.

The revenue raised by the tax is pQt, or the area of the rectangle bounded by $p(1+t)$, A, B and p. Since the revenue raised by the government falls short of the costs imposed on suppliers and consumers, the tax imposes *efficiency losses* equal to the 'triangular shaped' area AEB.

Diewert and Lawrence (1995) calculate that the efficiency losses imposed by the current taxes on labour income in New Zealand are in the order of 18% of the revenue raised, while for consumption taxation the losses approximate 14% of revenue. They suggest that the efficiency losses from the taxation of capital income are likely to be much higher than these figures, but they do not estimate them. The reason for the higher losses in the case of capital income is that the supply of investment funds (from abroad) is likely to be quite elastic.

It may be useful to relate this argument to the discussion on page 54 of Chapter 3 concerning the efficiency of private ownership. Recall that Epstein (1995) claimed the common law rules of ownership of productive resources and the income derived therefrom provided the most efficient incentives for decision-makers. Taxation can be thought of as imposing joint ownership of productive resources, with one of the owners essentially playing the role of a passive shareholder. Decision-makers face distorted incentives regarding how best to deploy the resources under their control.

The above analysis makes a further point, however. As Figure 4.1 illustrates, only part of the efficiency losses (area BDE) are borne by producers. In the case illustrated, consumers also suffer a share (area AED) of the loss. The value of the reduced consumption of market goods to consumers exceeds the share of tax revenue that they effectively pay. Taxation also distorts consumer incentives by making consumers think that the cost of satisfying their desires is higher than it really is.

In addition to the efficiency costs, there are costs of administering the tax system that are borne by the government, and compliance and avoidance costs that are borne by individuals (in terms of expenditures on accountants and lawyers, additional record keeping and so forth). These expenditures are also wasted since the resources involved could have been used to provide other valued goods or services. Diewert and Lawrence (1995) report that in New Zealand these wasted resources are at least 7% of revenue raised. Thus, leaving aside losses from the taxation of capital income, the best estimate of the total losses imposed by the current tax system in New Zealand is that they approximate 25% of the revenue raised. To produce a net gain in efficiency, therefore, marginal government expenditure has to deliver a benefit of about $1.25 for every dollar spent. Welfare expenditure also has substantial administrative and efficiency costs (the latter as a result of moral hazard). These would raise the required social benefits to far in excess of 25 cents for every dollar redistributed.

What type of compulsory insurance?

Compulsory redistribution of income through the tax and welfare system could be efficient if, despite the efficiency losses due to taxation and the distorted incentives of recipients, the anticipated insurance benefits exceed the expected costs. The high costs from distortions due to taxation and moral hazard on the part of recipients imply, however, that an efficient insurance scheme would be severely limited in scope. Knowing the future incentive effects of taxes and transfers, but not knowing their own particular future situations, rational taxpayers would choose a form of insurance from behind the veil of ignorance that concentrated on assisting those in desperate circumstances.

We observe a similar situation in private insurance markets. One way insurance firms control moral hazard is by using high deductibles. An insured individual bearing the full cost of minor expenses, has efficient incentives to economise on use and to find low cost service providers. Risk averse people are most concerned to protect themselves against catastrophic events. These events tend to be subject to fewer moral hazard problems, since the customer has less control over either their probability of occurrence or the costs incurred once the event happens. Insurance with a high deductible leaves the customer covered for the catastrophic events but avoids the efficiency losses associated with coverage of all events, large and small.

In the context of income redistribution, a high deductible insurance policy would cover catastrophic events that result in poverty, but leave private arrangements to cope with smaller fluctuations in annual incomes or expenditures. Redistribution from a year in which annual income is $50,000 to one in which it is $40,000 would yield too few benefits in the form of reduced variation in annual incomes to compensate for the efficiency losses of taxes and transfers. On the other hand, redistributing from a year in which annual income is $100,000 to one in which it is $0 may yield sufficient insurance benefits to compensate for the costs. Even in the latter case, however, private insurance may be available and, if so, it would be less costly than the tax and transfer system. Furthermore, private saving is an alternative form of self-insurance that does not suffer from moral hazard. More than likely, most people would prefer to rely on private saving, or borrowing against future income, to tide them over a small number of years of low income before they chose to incur the large efficiency losses associated with income taxation and income transfers.

Private insurance or savings, or compulsory government taxes and transfers, do not exhaust the possibilities for insuring against fluctuations in income or expenditure. Many people who find themselves in temporary difficulty receive income transfers from family members or friends. These private transfers may be subject to fewer moral hazard problems because the family members or friends may know much more about the individual's personal circumstances. The donors are therefore likely to be more knowledgeable about whether a recipient is truly in need, or has adjusted circumstances to take advantage of the opportunities that are available.

In summary, using income taxes and income transfers as anything other than an insurance scheme of last resort will be very expensive. Transfers ought, therefore, to be restricted to people in dire circumstances. The argument that governments should provide an insurance scheme of last resort does not imply that taxes and transfers ought to be directed at equalising incomes.

Legitimacy of the social order

People may voluntarily support income redistribution in order to maintain social stability. As labour economist Finis Welch stated in a recent paper on inequality:

> I would argue that inequality is destructive whenever the low-wage citizenry views society as unfair, when it views effort as not worthwhile, when upward mobility is viewed as impossible or as so unlikely that its pursuit is not worthwhile. Even more extreme, inequality can be literally destructive if it leads to "illegal" attempts to redistribute. These are issues that arise from the rules of the game, the role of law, freedom, and private property. (Welch, 1999: 2)

If the distribution of access to material goods and services is very skewed, as it is in some third world societies, internal law and order may become fragile. Poverty stricken citizens, who nevertheless are armed and healthy enough to mount a credible threat, may challenge the existing legal and political order. Redistribution might be seen as an alternative to police protection, or to personal expenditures on fences, alarms, guard dogs and private militia. Although this example is not relevant to New Zealand, it may be of some value to discuss it briefly.

Redistribution motivated by a concern for social stability or fear of a rebellious underclass is no less susceptible to free rider problems than is redistribution based on compassion. Suppose that most of the material wealth in a country is controlled by a small number of families. While all of the wealthy families may benefit from maintaining the existing economic, political and legal institutions, each family may have an incentive to allow the other families to do most of the redistributing. Any one family may then consider only the private costs and benefits of redistributing income. The result could be that the total amount redistributed is inefficiently small. In particular, each family might agree to be forced to contribute taxes *ex post* if all the other families also agree to accept the same coercion. Even though any one family would have to make higher contributions as a result of such an agreement, the additional benefits from having other families simultaneously contribute greater amounts of tax could more than offset the increased private costs.

Marginal versus infra-marginal benefits

An important point to keep in mind when considering whether the amount of private redistribution is likely to be efficient is that only the costs and benefits of an additional dollar on income transferred are relevant for efficiency. Even if the total benefits of income transfers exceed the total costs, the existing amount of redistribution could be efficient. For inefficiency, the marginal benefits of an additional transfer must differ from the marginal costs.

For example, suppose many people expect their life will be better if they live in a community in which everyone can read, write, do basic arithmetic and demonstrate a familiarity with the economic, legal, and political institutions. Obtaining a basic level of education that imparts these skills would then confer benefits on other individuals.

In the absence of any subsidies to undertake education, any one family would consider only its own benefits and costs when deciding how much education to purchase for its children. The benefits to non-family members of providing children with a basic level of education would not influence parental decisions. It does not necessarily follow, however, that the level of education that people would choose to undertake would be inefficient. Suppose that the private benefits of education for any individual are so high relative to the costs that everyone would purchase more than the basic level in any case. The benefits to others of additional education beyond this basic level could be zero. Even though the basic education is providing spill-over benefits to others, the voluntary outcome would be efficient.

The point can be illustrated with the aid of a simple diagram. In Figure 4.2, the curve MB_P represents the marginal private benefits of additional education to the family making the decision. All remaining individuals in the economy have the relatively low marginal benefit MB_s from additional education of the children in this family. Since there are a large number of these individuals, however, the total marginal benefits of additional education for all of them as a group, or the vertical sum of their marginal benefit curves, will be very large for low levels of education but also very steep. The marginal cost of additional education is represented by the upward sloping curve MC.

If the choice of the level of education is left entirely to the family, family members would choose the level at which the private marginal benefit matches the marginal cost of provision. At this level, the *marginal* benefits for the myriad remaining consumers is zero. Even though the family is

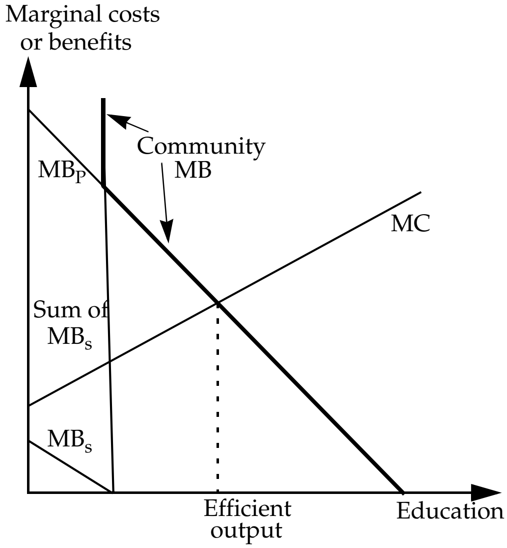

FIGURE 4.2: An efficient outcome despite spill-over benefits

providing spill-over benefits to everyone else in the economy, it would not be efficient to subsidise its consumption of additional education.

Redistribution in cash or in-kind

The discussion of subsidising educational expenditure raises another efficiency issue. Certain types of expenditures, such as nutritious food and basic or minimal levels of shelter, education, health care or legal representation, are often thought to be more worthy of support than are other expenditures. These types of goods or services are regarded by many as being essential to giving people a fair opportunity to succeed. At the opposite extreme, expenditures on illegal drugs, gambling, junk foods and so forth are considered undesirable and not worthy of support.

Economists have generally argued against subsidising the consumption of particular goods and services and instead have suggested that income transfers are preferable. Subsidies tied to particular expenditures can at best benefit the recipients by as much as an untied income transfer of equal cost to the donor. A recipient who desired to spend the income on the targeted services could always do so. A recipient who chooses not to spend the money in that way presumably believes that the alternative purchases are better.

A contrary argument is that the utility gains to a donor from seeing expenditures the donor regards as worthwhile, and the utility costs of seeing transfers wasted or worse, are also part of the utilitarian calculations that underlie efficiency. The welfare of donors as well as recipients enters calculations of the "greatest happiness of the greatest number". As we suggested in Chapter 2, a concern for individual freedom may better explain the usual preference for letting recipients decide how to spend their income.

While there is a general presumption that in-kind transfers are likely to deliver fewer benefits to the recipients than income transfers, in-kind transfers may be justified in extreme cases. Just as the state intervenes to take children away from abusive parents, there may be situations in which the consumer sovereignty of parents, and their right to make choices on behalf of their children, ought to be overruled. Such interventions ought, however, to be rare.

In-kind transfers may suffer from other defects that make them less desirable than income transfers. While the consumption of particular goods or services could in theory be subsidised with monetary payments tied to expenditures, in practice the subsidised services tend to be supplied by public agencies. The result is that the goods or services are often not supplied at the lowest possible cost, or at the highest quality, given the costs that are incurred. We argued in Chapter 3 that this may reflect, in part, a desire to serve the self-interest of suppliers. The privileged suppliers in turn become powerful advocates for continuing in-kind subsidies as opposed to untied income transfers. Indeed, it is perhaps not surprising that services supplied by public agencies are routinely touted as providing external benefits, while it is assumed the taxed activities do not provide such benefits to a comparable extent, even if we do not know exactly what the taxed activities are.

Services provided by public agencies, such as public education, are also often made available to everyone regardless of the level of need. Suppliers are likely to favour this arrangement, since they will most likely benefit from a higher level of expenditure on the subsidised service.

For example, if the money spent on public education were left in the hands of taxpayers, in general only part of the funds would be spent on education services. Typically, the funds would be spread across all the goods and services that consumers value.

On the other hand, the taxes imposed to pay for public education will engender efficiency losses. These will appear as a lower level of utility for consumers in a world in which taxes are higher. The lower utility is likely to take the form of reduced total expenditure on marketed goods and services. Higher taxes on market activity encourage households to substitute untaxed activities, such as leisure pursuits, for the consumption of marketed goods and services. If total expenditure is higher in the absence of the taxes imposed to finance public education, some of this increased expenditure may raise the private demand for education. Nevertheless, empirical evidence on the size of the disincentive effects of taxes and incentive effects of subsidies strongly suggests that the overall expenditure on educational services will be higher under a system of public education provided at no charge to consumers.

Finally, the excessive cost of public education relative to private alternatives is likely to imply that more private education services could be purchased for the same total expenditure on education. Thus, the output of quality-adjusted educational services could be higher under private provision even if the expenditure on education is lower. This outcome is unlikely to be desirable from the perspective of teachers, school administrators or other suppliers, however, since they are the major beneficiaries of the excessive costs in public education.

Whether or not the total demand for educational services is higher or lower under public provision, suppliers are likely to enjoy another major benefit of universal access without user charges. The publicly supplied product, without any user fees, will be strictly preferred to the private alternative by most customers. This is likely to allow the suppliers to extract some rents. The rents could take the form of excessive wages or uncompetitive working conditions (if the suppliers can control entry to the business), or perhaps a quality of service the suppliers would prefer to supply rather than a quality of service the customers would prefer to consume.

In general, in-kind transfers therefore are likely to be much less efficient than the cash alternatives. If in-kind transfers are used, however, they ought to take the form of a subsidy to competitive private supply rather than being provided by a monopoly government supplier. The consumption subsidies, or vouchers, ought also to be restricted on the basis of need, in order to avoid the efficiency costs of universal access.

IS ANYTHING BETTER ACHIEVABLE?

There is another point to keep in mind when we are considering the efficiency of financial and insurance markets and private income transfers. Although voluntary arrangements may not result in the most efficient allocation of resources imaginable, such arrangements may nevertheless produce the most efficient outcome that is obtainable in practice. There may be no way of altering the private equilibrium to secure the imagined efficiency gains.

In particular, giving someone the right to compel others to transfer income opens up the possibility of imposing transfers that go beyond anything justifiable on efficiency grounds. Excessive income transfers obtained by coercion could be much more inefficient than a system of personal savings, insurance and private transfers.

Voluntary private arrangements are likely to result in some income redistribution. Many religious, cultural, educational and medical institutions, for example, have been financed largely through private charity. Many employers also provide charitable benefits for their employees, such as assistance with further education and donations made to charities supported by employees. In small and remote communities with one major local employer (such as a mining firm) many public goods are provided by that employer. In addition, many individuals contribute time and other non-monetary resources to charitable causes. Various clubs and other voluntary organisations supply many types of public goods at the local level.

Social pressure is effective at overcoming free rider problems in many of these situations, especially in smaller communities with a stable population. People know those who are contributing to local joint activities and those who tend to hold back and wait for others to take the initiative. There are many ways of rewarding prominent public-spirited citizens, such as presenting certificates or making public pronouncements, that are not costly. Such measures may, however, be effective at encouraging voluntary provision of jointly consumed public goods.

Even if compulsion can increase the provision of public goods, it nevertheless has other undesirable consequences. When people are compelled to do something, they will tend to resist. There will be an inevitable erosion in community cohesion. People may also withhold information about the benefits they obtain from publicly provided goods and services if the consequence is that they will be called upon to make a greater contribution. On the other hand, if contributions are unrelated to the benefits derived from the joint service, people paying less than the average level of taxes will have the opposite incentive to overstate the value they place on the public activity. In either case, the government is likely to have information about the benefits of public services that is inferior to the information revealed through people's behaviour in making voluntary contributions. The costs of providing joint services may also be distorted when people can be compelled to contribute resources without having to be compensated for the lost opportunities to use the resources for other purposes. When the government is allocating resources on the basis of inaccurate information about both the benefits and the costs, it is hard to imagine it will do a very good job of allocating resources efficiently.

Finally, there is evidence that compulsory contributions to jointly provided services tend to displace private contributions. Public provision of many services such as education and income support has also tended to weaken the family ties that are important in ensuring intergenerational mobility. These issues are discussed in more detail in Chapter 6.

SUMMARY

Efficiency and equality are often seen as conflicting goals, although this need not be the case. Some policies and institutions can be seen as furthering the goals of efficiency and vertical equity, but only when the latter is viewed as a compassionate concern for people living in poverty. Redistribution based on compassion may be regarded as a means to the end of assisting efficient resource allocation in a world of risk and uncertainty. In Chapter 5 we consider whether equality is valuable in and of itself, rather than as a means to other ends.

5
WHAT'S WRONG WITH EQUALITY?

In Chapter 4 we argued that limited redistribution aimed at alleviating poverty may be consistent with promoting efficiency. More extensive redistribution is likely, however, to come at a high cost in terms of lost efficiency. Society faces a trade-off between policies that increase average standards of living and policies aimed at equalising income or wealth. Over the medium term, a higher rate of economic growth does more to raise the standard of living of people on low income than does redistribution accompanied by a lower rate of economic growth:

> Wherever the free market has been permitted to operate … the ordinary man has been able to attain levels of living never dreamed of before … Industrial progress, mechanical improvement, all of the great wonders of the modern era have meant relatively little to the wealthy. The rich in Ancient Greece would have benefited hardly at all from modern plumbing: running servants replaced running water. Television and radio – the patricians of Rome could enjoy the leading musicians and actors in their home, could have the leading artists as domestic retainers. Ready-to-wear clothing, supermarkets – all these and many other modern developments would have added little to their life. They would have welcomed the improvements in transportation and medicine, but for the rest, the great achievements of Western capitalism have redounded primarily to the benefit of the ordinary person. (Friedman and Friedman, 1979: 146–147)

Economic growth and technical progress have raised standards of living not only by increasing the quantity and quality of consumer goods but also by reducing the burdens of work. A rational individual behind the veil of ignorance would sacrifice efficiency and economic growth in order to transfer income to families who are not in dire need only if that individual considered equality to be quite valuable in itself.

What is wrong with seeking equality of outcome? Utilitarians, Marxists and others have offered arguments purporting to show that the only fair distribution of market goods and services is an equal one. Some modern scholars argue that equality of material wealth has value in and of itself, and not merely because it contributes to some other end. Temkin (1993: 7) notes:

> Although there may be many differences between them [egalitarians], as I use the term an *egalitarian* is anyone who attaches some *value* to equality *itself*. That is, an

> egalitarian is anyone who cares *at all* about equality *over and above the extent it promotes other ideals* [emphasis in original].

While scholars such as Temkin argue for the value of equality of market income or material wealth for its own sake, others claim that equality of income or material wealth, or at least some redistribution of income, will help achieve other desirable goals such as efficiency. Some of these arguments were discussed in Chapter 4.

Temkin's position appears to countenance envy as an acceptable social doctrine. He suggests that a state of perfect equality *E* (where everyone has identical wealth *W*) would be better than state *I*, a state which is identical to state *E*, except that in state *I*, person *X* has wealth *W* + 1. Person *X* is better off in state *I*, and no one is worse off (with regard to *W*) in *I* than in *E*.

Temkin remarks:

> I realize most economists think it conceptually true that if one outcome is pareto superior to another then it must be better all things considered. But … I am not convinced of this. In particular, it is arguable – though controversial – that egalitarians should maintain that in *some* cases pareto improvements worsen a situation all things considered. Consider, for example, the effect of raising one person a certain amount in an otherwise perfectly equal world. It is by no means clear – to me anyway – that the gains in utility and perfection accruing to that one person must necessarily outweigh the egalitarian complaints of those left behind. There might be ten billion left behind. Even if each of their complaints would be relatively small, together they might outweigh the gains of the one person benefited. I think it is at least an open question whether the egalitarian objection to certain pareto improvements might be sufficiently strong as to condemn those improvements all things considered. (Temkin, 1993: 139–140)

In a recent essay in *Public Interest*, Feldstein (1999: 33) states the contrary position:

> Imagine the following: Later today, a small magic bird appears and gives each *Public Interest* subscriber $1,000. We would all think that this is a good thing. And yet, since *Public Interest* subscribers undoubtedly have above average incomes, that would increase inequality in the nation. I think it would be wrong to consider those $1,000 windfalls morally suspect.

Similarly, in a recent letter to the editor of *The Economist* magazine, economist Hausman (1999) gave the following example of a poor public policy decision that could result from an exclusive concern about inequality:

> Suppose that Bill Gates wakes up one morning and decides that America's Department of Justice has gone too far. He decides to move Microsoft to … Britain, and you might expect any rational government to welcome Microsoft with open arms – no pollution, high-paid jobs, significant tax payments with little

> additional requests for excessive welfare payments and no requests for state subsidies or tax breaks as with the relocation of most other transnational corporations.
>
> However, you would be wrong, because a young academic sociologist at the London School of Economics will soon point out that the move would have a potentially devastating effect on the government's favourite poverty measure.

Pareto improvements that make only people who are wealthier than average better off have a negative effect on measures of inequality. An exclusive focus on inequality therefore leads one to reject such improvements even though they directly harm no one.

Economists such as Sen (1997), who believe that reducing inequality is valuable, nevertheless recognise also that efficiency is desirable and increases in efficiency can more than compensate for increases in inequality. As we note later in this chapter, the social welfare function is the most widely accepted framework among economists for formalising a trade-off between efficiency and equality as potentially competing goals. Under this approach, a Pareto improving reallocation of resources will always be valued positively even if only the wealthiest people are made better off. Other widely used measures of income inequality that we also shall discuss, including the Gini coefficient and the coefficient of variation, do not share this property. Economists who advocate using the Gini coefficient and other simple measures of inequality usually also suggest that such measures need to be supplemented by efficiency measures. An increase in efficiency could then compensate for an increase in inequality and lead to a favourable overall assessment of a Pareto improving reallocation of resources.

The main message of this chapter, however, is that society ought not to pursue equality of income or material wealth, either as a means to an end or as an end in itself. At best, equality of material wealth or income are proxy goals. If we are concerned about people being unequal, what should concern us is inequality of outcomes, as people perceive them. Why should we fixate on a subset of the items that determine the value people place on different states of the world?

As noted in Chapter 2, the common sense principle is the most basic principle of ethics. The goal of equality of outcome violates this principle because it is impossible to say exactly what the goal is, or even to measure progress toward achieving it. The goal attains some intuitive plausibility only because it is related to, and is confused with, equality of measured income or material wealth.

Suppose we ignore the substantial problems associated with approximating outcomes by access to income or material wealth. The common measures of inequality in income or wealth nevertheless are seriously flawed, in so far as they are predicated on the notion that envy is socially acceptable. As Feldstein (1999) argues: "What policy should address is not inequality but poverty".

'EQUALITY' – AN ILLUSORY GOAL

What does 'promoting equality' mean? An analogy can be made with dividing up a cake. The fairest way of dividing a cake is by cutting it into equal slices for everyone. To what in society, however, does the cake correspond? Is it money, happiness, material goods, experiences? Does 'promoting equality' mean that everyone should have exactly the same material assets, exactly the same consumption bundles, exactly the same money incomes or an equal share of something else? As Sowell (1999: 52–53) has recently remarked:

> One of the reasons why equality may be impossible to achieve is that merely defining it opens up a bottomless pit of complications … equality among multidimensional entities, even inanimate and abstract entities like the Gross National Product, may be impossible to define – much less achieve.
>
> In the case of determining which country has a higher Gross National Product, the first country may have a higher GNP if the common denominator is the price of goods as measured by using official exchange rates, while the second country may have a higher GNP if the test is which country could actually purchase all of the other country's GNP in the international marketplace and still have something left over …
>
> Since human beings are even more muti-dimensional, defining equality among them becomes even more problematical and ultimately arbitrary.

Even in ancient times, there were prominent politicians and philosophers, such as Plato, who favoured the equalisation of material goods, of radical redistribution from the wealthy to the poor. Aristotle pointed out some of the problems with this position.

> Plato in the *Laws* was of the opinion that, to a certain extent, accumulation should be allowed, forbidding, as I have already observed, any citizen to possess more than five times the minimum qualification. But those who make such laws should remember what they are apt to forget – that the legislator who fixes the amount of property should also fix the number of children; for, if the children are too many for the property, the law must be broken …
>
> Again, where there is equality of property, the amount may be either too large or too small, and the possessor may be living either in luxury or penury. Clearly, then, the legislator ought not only to aim at the equalization of properties, but at moderation in their amount.

> And yet, if he prescribe this moderate amount equally to all, he will be no nearer the mark; for it is not the possessions but the desires of mankind which require to be equalised … (Morris, 1959: 25)

As Aristotle noted, equalising the amount of money people possess will not equalise their well-being. Large families require more money than smaller families if they are to achieve the same material quality of life (although children may provide non-material benefits to their parents). Sowell (1999: 67–68) also points out the number of income earners in a family will affect measured family incomes:

> There are more than 50 percent more people per household earning $75,000 and up as per household earning under $15,000. That is one of the reasons for their differences in income: People earn money and more people tend to earn more money.
>
> There are more than twice as many income-earners per household in households earning $75,000 and up as per household earning under $15,000. The top 20 percent of income-earning families supply 29 percent of all workers who work 50 weeks or more per year, while the bottom 20 percent supply just 7 percent of such workers.

Equality of money incomes or material wealth

At first glance, it seems obvious that promoting equality of money incomes is a feasible and appropriate goal. If we want to make people more equal, we take money from those with a higher income and give it to people with a lower income. Income taxes and welfare payments of various sorts can be seen as an attempt to promote after-tax and after-transfer incomes that are more equal than they would otherwise be.

As we noted when discussing the ownership of market incomes in Chapter 3, however, people care about much more than money income. People's preferences can have very complicated effects on our ability to draw inferences from the economic statistics we collect. Several examples may help clarify this point.

Most people earn money income not as an end in itself but rather for the marketed goods and services that money can buy. The purchasing power of a given money income depends on how the money is spent. The same money income will buy far fewer goods and services when those goods or services have a high price. This fact is routinely noted in international comparisons of 'standards of living'. It does not make sense to compare the dollar incomes of a New Zealand resident with the New Zealand dollar-equivalent income of a villager in India. Converting incomes measured in rupees into New Zealand dollars, and comparing them with New Zealand incomes, gives a poor idea of the relative standards of living in the two countries. The prices of many services and goods differ by more than the exchange rate between the two currencies. Instead, economists attempt to make purchasing power corrections by examining the bundles of goods and services that incomes of different levels can purchase in either country.

Similarly, a daily commuter working at an office job in Auckland needs a lot of money for clothes, travelling expenses, for the high costs of living in a popular urban area and so on. The same money income will buy a substantially different bundle of goods or services in a small town or a rural area. Furthermore, the rural resident may have access to some goods or services, such as fresh air, that the urban resident cannot buy at any price. Conversely, the urban resident will have access to types of entertainment and other services that may be available to a rural resident only at a very high price, if at all.

Indeed, many people have the opportunity to choose where they will work and live, and one component of that choice is the money income they can earn. Many happily forgo the opportunity to earn a higher money income if, in return, they are able to live near friends or family, and enjoy nature, a more friendly community and a less hectic lifestyle. Upon what basis do we conclude that such people have made a mistake? They apparently do not believe that their lower money income makes them worse off. If they valued a higher money income more than a pleasant lifestyle, they would have chosen the higher money income. Taking money income from urban commuters to give to their rural cousins may then increase inequality in standards of living, even while it reduces an inequality in money incomes.

There are many other ways people can trade-off money incomes for other things they value. Different types of jobs have different rewards or satisfaction associated with them. People will happily forgo some money income in return for more control over their working conditions, or the amount of say they have in the way the business is operated. Some people place a high value on flexibility in working hours, or the opportunity to take many vacations at times of their choosing. Some value a pleasant working environment, or congenial workmates or bosses, and are willing to sacrifice money income in order to obtain those things. Conversely, some jobs may involve high levels of physical danger, or a tolerance for heights or unpleasant smells that most people would prefer to avoid. Higher salaries must accompany these unattractive jobs before people voluntarily choose to take them. In any of these situations, redistributing from people with higher money incomes to people with lower money incomes can exacerbate inequality in standards of living, even while reducing inequalities in money income.

Jobs differ in the number of hours of work needed to earn a given money income. Many people can earn income by exploiting attributes, such as musical, sporting or mathematical ability, that are not themselves equally distributed in the community. The benefits do not accrue, however, merely from the possession of the required genetic traits. Talents need to be nurtured and developed in order to generate useful income. Yet how could one adjust for the benefits conferred by these genetic advantages? How much of the observed income differential between Luciano Pavarotti and us is the result of his natural genetic endowment and how much the result of choices he has made (such as hours of practice and a willingness to travel extensively) that may have reduced his standard of living in other ways? Even if one could, in principle, adjust for some of these factors, how expensive would it be to do it in practice? The efficiency costs resulting from the disincentive effects of both taxes and transfers would add to these administrative costs of redistribution. The resulting costs are likely to outweigh any of the supposed benefits arising from the greater equality that may result.

Even if we suppose that appropriately adjusted incomes could be equalised, we would not achieve the ostensible goal of equal living standards. Households differ in their ability to convert money incomes into effective consumption. For example, there is evidence that education allows one to become a more effective consumer of goods and services. A well-educated person is often better able to assess the qualities of service providers, may find it easier to search for alternative sources of information on products, and may have a better foundation for obtaining value from services, such as listening to classical music on the radio or visiting art galleries, that may not cost a lot in money terms.

Another major issue is the appropriate definition of the income earning, or consuming, unit. Should we focus on individuals, nuclear families or extended families? Most children have a very low measured income, yet it is ridiculous to say that they are worse off than most adults, regardless of the resources their parents make available to them, or allocate on their behalf. (Creedy (1997: 29–31) discusses the effect on measured inequality of different assumptions about income sharing within households.)

In many households, an adult voluntarily withdraws from the labour market to raise children. The value of the children and the home services are judged to be more than compensation for the reduction in market income. Some might even say that adequate childcare outside the home cannot be purchased at any price. The only market substitute may be a live-in caregiver who looks after a child from birth and comes to care about the welfare of that child in the way a stay-at-home parent would.

People also provide many other home services that are not recorded as market income. Families in which someone cooks the meals, mows the lawn, makes home repairs, washes the clothes, cleans the house and so forth may have a higher standard of living than otherwise comparable families who purchase all or most of these services on the market. Creedy (1997: 32) suggests that time-use surveys might be used to adjust income data for home production, but little appears to have been accomplished in practice.

The appropriate definition of an income-earning unit can also extend beyond a nuclear family. Even when nuclear families live apart, some brothers, sisters and parents share income and other resources (such as childcare services) with extended family members. The same measured household income can, under those circumstances, provide very different standards of living.

Redistribution programmes implicitly assume that market income is properly attributed to individuals. When income can be channelled through entities such as family trusts and private companies, however, reported personal incomes may provide a very poor indication of the real income earned by a person.

How do we adjust for disabilities, or severe medical conditions? For example, if we were to adjust money incomes to arrive at standards of living, we would need to adjust for large medical expenditures. Someone who earns $100,000 in a given year but has to spend $80,000 on medical treatment to obtain a lifestyle that is close to normal is probably worse off than someone else earning only $20,000 but who is in good health and has no large medical expenses. We may also ask whether it is sufficient to adjust incomes for explicit payments for hospitals, doctors, drugs and so on. Might there be additional psychological costs associated with severe medical conditions?

Many of these corrections that ought to be made to money incomes in order to estimate standards of living share an awkward feature. There is no objective way of measuring them. When someone chooses one option over another, it is reasonable to presume that the chosen option is preferred. But different people might choose differently, and might value consumption and employment opportunities differently. Is one valuation objectively better than another? Is there just one appropriate way to adjust money incomes to arrive at measures of standards of living?

Suppose that we could adjust money incomes for differences in the non-pecuniary aspects of different jobs and for differences in consumption bundles. We would still not be able to pursue a goal of greater equality. Money incomes in any given year, adjusted or otherwise, are not necessarily representative of what a particular person could expect to earn in a normal year.

For example, students attending university may have very low money incomes, and very low levels of expenditure on personal consumption items. Most of these students are destined, however, to enter the ranks of the highest money income segment of the population as future doctors, lawyers, accountants, engineers and other professionals. Redistributing money income from a building worker to a university student may reduce inequality of money incomes in a given year, but it will almost surely increase inequality of lifetime earnings.

Some of the adjustments necessary in theory would be impossible to make on objective grounds in practice. People make different trade-offs between current and future consumption. Some are willing to earn a low income for some time in order to earn a much higher income after completing additional education or training. Others find the current sacrifice in earnings insufficient to compensate for the future increase. It is also difficult to correct for potential differences in lifespans. If we take too much (relative to a theoretically optimal level of redistribution) from someone who dies at an unexpectedly young age, there is no way of reimbursing that person once the excessive redistribution has been revealed.

As with incomes, expenditures in any one year can be unrepresentative. We have already observed in Chapter 4 that financial markets allow individuals to spread their consumption over time, to self-insure against fluctuations in incomes and expenditures. This is known as the *permanent income* or *lifecycle* theory of consumption. Economists have shown that people tend to base their expenditures on their permanent, or normally expected, levels of income than on their actual measured incomes in any given year. In addition, people tend to pass through lifecycles in which they borrow against expected future income while establishing households and developing careers, then save for retirement at the ends of their working lives:

> A major factor in both income and wealth is age. Those who have worked for many years tend to have advanced in their careers to higher-paying positions and to have accumulated more assets, whether in the form of money in the bank or a pension fund, or equity in their homes. People in their sixties have persistently

> had higher incomes than people in their twenties and much higher net worths. Sowell (1999: 57)

Basing income redistribution on the expenditure, income or wealth of an individual in any given year could increase inequality of lifetime welfare levels.

Redistribution programmes often transfer income from one stage of life to another. For example, redistribution programmes increase income when a person is young but reduce it in middle life. Along the way the large efficiency losses of taxes and transfers mean that the present value of receipts falls short of the present value of payments. Such intertemporal transfers therefore become an expensive substitute for individual saving.

Many people who are impoverished remain so for only a brief time. While 'the poor' may always be with us, not all the current people who are poor will remain that way until the ends of their lives, or even for the next five years. People often end up in poverty following a temporary cataclysmic event, such as divorce, serious illness or unexpected unemployment. Once they have had time to adjust their lifestyles, find new places to live, find new jobs, or call on assistance from friends and family, people can often recover from such disasters to achieve standards of living approximating the ones they had previously enjoyed. As Sowell (1999: 54) observes for the United States:

> People in the bottom quintile of the income distribution spend nearly two dollars for every dollar of income they receive. Two-thirds of the statistically defined "poor" have air-conditioning, and more than half own a car or truck. More than one-fourth of "the poor" own two cars and/or trucks and hundreds of thousands of them own homes costing more than $150,000. Puzzling as such anomalies might seem if we were discussing an enduring class of genuinely poor people, they are understandable in a statistical category which includes many transients. In any given year, many entrepreneurs may be earning not only low incomes but negative incomes as their businesses incur losses. Professionals, entertainers, and others may also suffer off years that leave them in the bottom 20 percent for that year. Many members of high school and college graduating classes enter the labour force in the middle of the year, earning only about half of what they will normally be earning when they work the entire year.

Redistribution programmes are misguided if they focus on temporary difficulties that affect many people. We may end up taking from someone who has a low, but stable, standard of living in order to assist someone who is temporarily disadvantaged, but who actually experiences a much higher standard of living on average.

Equality of 'utility'

Although we can measure income, expenditure and material wealth, we cannot deduce anything about a person's level of happiness or 'utility' from these measurements. 'Utility' is a technical term used to describe units of happiness, pleasure or goodness, as perceived by each individual. A person is said to be maximising utility when actions are taken, within the constraints imposed by limited available resources, to increase well-being as measured by that person's own tastes or preferences.

The fundamental problem with measuring utility is that observed behaviour always involves the choice of one item or activity versus another. A dollar of income can be spent on either item *A* or item *B*. An hour of time can be spent in activity 1 or activity 2. One's limited energy or powers of concentration can be allocated to one end or another.

In each case, maximising behaviour implies that equal additional utility should be obtained from using more of one of these limited resources in any one use. For example, the additional utility obtained from the last dollar of income spent in each possible way should be the same. If this were not the case, less of the resource in question could be used where it was less able to generate utility, freeing up more to be used where it is more productive in generating additional utility. For someone who is maximising, the additional benefits obtained from additional use of each limited resource in each activity will be equal. We would observe a change in choice behaviour if, and only if, this equality condition were not satisfied.

The key result, however, is that the equality condition is unaffected by a rescaling of utility. Any such rescaling would affect all the benefits in the same way, and the rescaling factor would cancel out from all the equality conditions. Any use of limited resources that maximises utility would continue to maximise no matter how utility is rescaled.

Technically, utility is an ordinal, and not a cardinal, concept. Temperature provides a physical analogy. We can say that one temperature is higher than another so that temperatures can be ordered. But it does not really matter whether we measure temperatures on the Fahrenheit or the Centigrade scale, as long as we are consistent. If two temperatures have the same value on one scale, they will also have the same value on the other scale.

The ordinal nature of utility reflects the fact that observed behaviour reveals a person's relative valuations of different choices, but tells us nothing about the absolute level of utility that the person obtains in any situation. Furthermore, the equality condition pertains only to the last unit of limited resources allocated to each alternative. Someone who is unwilling to spend time, income or energy obtaining an additional sip of water nevertheless is likely to value the water that has already been consumed. All we can deduce from an unwillingness to incur additional costs is that the marginal value of water is zero, given what else is already being consumed or could be consumed in the future. Since one cannot live without safe drinking water, the total value of drinking water is undoubtedly large, even if the incremental value of water declines rapidly as more is consumed.

Knowing that people equalise utilities at the margins tells us nothing about the level of utility that they gain from infra-marginal expenditures. Thus, even if we could somehow show that two people obtained identical marginal utilities from a given unit of wealth, they could still have very different total utilities.

In summary, if the goal of equality means anything it has to mean something similar to equality of utility. Measures of income or expenditure need to be adjusted in order to arrive at measures of standard of living. Such adjustments, however, require unambiguous measures of total utility when, at best, objective evidence only tells us about ordinal measures of marginal utility. In addition, of course, an enormous amount of information would be needed to make such adjustments, even if they were possible in principle.

In practice, redistribution has to rely on measurable quantities such as monetary income or wealth that are a crude approximation of the theoretical concept of interest. Such measures are also easy to manipulate in order to increase eligibility requirements. Only the most extreme differences in standards of living could be reduced by relying on such measures.

MEASURES OF INEQUALITY

Conceptual and practical difficulties have not prevented scholars from proposing measures of inequality. Most of these measures simply ignore the conceptual and practical difficulties and focus on what can be measured, that is, market incomes or material wealth.

In the old-style welfare economics first proposed by writers such as Jeremy Bentham and John Stuart Mill, market income was assumed to affect each person's utility in exactly the same way. The primary distinction between 'old' and 'new' welfare economics is that utility was taken to be a cardinal concept in the former, while it is an ordinal concept in the latter.

Old welfare economists further assumed that, for higher levels of current income, a fixed additional increase in market income produced a smaller gain in utility. For example, an additional dollar of market income was assumed to give less additional utility to someone already earning $50,000 than to someone already earning only $20,000. As shown in the appendix to Chapter 3, the assumption of decreasing marginal utility of income or wealth corresponds to the empirical observation that people are risk averse.

Suppose that one assumes that the utility of one person is neither more nor less valuable than the utility of any other person, and (quite unrealistically) that redistribution can be achieved at no cost now or in the future. The total utility of the society as a whole would then be maximised when the distribution of market income is exactly equal. If income were not equally distributed, the total utility of the society as a whole could be increased by redistributing from a high income earner to anyone earning a lower income. The loss in utility from taking a dollar from the highest income earner would be less than the gain in utility from giving that same dollar to the lowest income earner.

The result that total utility could be increased by redistributing from a high income earner to anyone earning a lower income gave old welfare economists an egalitarian reputation. As Sen (1997: 16–18) observes, however, the egalitarian result is quite fragile. He gives an example of two people for whom the utility of any level of income for Happy is twice the utility that the same income would produce for Sad. If the incomes of Happy and Sad were equal, Happy would be enjoying twice the utility of Sad. The marginal utility of an extra dollar for Happy, however, would also be double the marginal utility of an extra dollar for Sad. Making the (unrealistic) assumption that redistribution is costless, an adherent of old welfare economics would, in this instance, recommend taking income from Sad and giving it to Happy in order to raise total social utility. Such a redistribution would not be egalitarian in effect, however, since it would increase the difference in utility between the two people.

Sen focuses instead on what he calls the weak equity axiom (WEA):

> *The Weak Equity Axiom:* Let person *i* have a lower level of welfare than person *j* for each level of individual income. Then in distributing a given total of income among *n* individuals including *i* and *j*, the optimal solution must give *i* a higher level of income than *j*. (Sen, 1997: 18)

Sen observes that the WEA can in principle allow for adjustments to income of the type discussed in the previous section in order to produce rough measures of utility:

> The framework in terms of which WEA seems to me to make a great deal of sense is the one that is being used in this work, viz., considering the possibility of being in different persons' positions and then choosing among them. Thus interpreted WEA amounts to saying that if I feel that for any given level of income I would

> prefer to be in the position of person *A* (with his tastes and his other non-income characteristics) than in that of person *B,* then I should recommend that *B* should get a higher income level than *A*. (Sen, 1997: 19)

This assertion simply ignores the difficulties of objectively measuring the "tastes and … other non-income characteristics" of other people in so far as their implications for utility are concerned. Furthermore, most of the discussion in the remainder of Sen (1997), as well as most of the discussion in the professional literature, focuses on inequality in measured money incomes. The implicit assumption of a perfect correlation between money incomes and utility or standards of living is rarely stated explicitly, let alone defended.

It may nevertheless be valuable to discuss some of the various measures of inequality that have been proposed. Let us suppose, for the sake of argument, that money incomes correlate perfectly with standards of living. Under these circumstances, the proposed measures reveal implicit ethical axioms about what one ought to like or dislike about different distributions of income. We argue that, when stated explicitly, those implicit axioms often have one or more unattractive characteristics.

The coefficient of variation and the Gini coefficient are the most commonly used measures of income inequality. Statistics New Zealand (1999) uses the Gini coefficient as the measure of inequality throughout its report. Supplementary statistics, including differences in average income by sex, ethnic group and educational qualifications are also presented, but not as measures of inequality. There is also a brief discussion (Statistics New Zealand, 1999: 74–78) of changes in the proportion of households that fall within given bands around median "equivalent household disposable incomes" (that is, incomes after adjusting for taxes, transfers and household composition). Chapter 6 of Statistics New Zealand (1999) examines some characteristics of the households with the lowest 20% of equivalent household disposable incomes, but explicitly disavows using such income standards to measure inequality. For reasons we shall make clear later in this chapter, we believe a 'poverty line' concept is the only relevant measure of inequality.

The simplest measure of inequality to explain, apart from the number of people below an income standard or poverty line, is the coefficient of variation. It is based on the standard deviation of the distribution of incomes. The explicit formulae are useful for understanding what this concept is measuring.

Suppose we have n people with incomes y_1, y_2, …, y_n. The average, or *sample mean*, income is the sum of the incomes divided by the number of them:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i . \qquad \text{(EQ 5.1)}$$

The *sample variance* of the distribution of incomes is the average of the squared deviations of each income from the sample mean:

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 . \qquad \text{(EQ 5.2)}$$

The *standard deviation* is defined as the square root of the sample variance, while the *coefficient of variation* is the ratio of the standard deviation to the mean:

$$CV = \frac{\sigma}{\bar{y}} = \sqrt{\frac{\sigma^2}{\bar{y}^2}} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{y_i - \bar{y}}{\bar{y}}\right)^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{y_i}{\bar{y}} - 1\right)^2} . \qquad \text{(EQ 5.3)}$$

The coefficient of variation thus measures the proportional deviation of incomes about the average income level.

We can illustrate these calculations, and other calculations in this section, using an imaginary economy with five people earning 0, 2, 5, 7 and 11. The mean income is then (0+2+5+7+11)/5 = 5, the variance is

$$\sigma^2 = \frac{(0-5)^2 + (1-5)^2 + (5-5)^2 + (7-5)^2 + (11-5)^2}{4} = 20.25 \qquad \text{(EQ 5.4)}$$

and the coefficient of variation is $\sqrt{(20.25)}/5 = 0.9$.

Intuitively, the amount of 'deviation' in a set of numbers is a measure of their dispersion both above and below their average. By adding the squared differences of each number in the set from their average, the sum is increased whether the particular number is above or below that average. To correct the units, the square root of the resulting 'average squared deviation' is taken. This is the standard deviation. But we still have not fully corrected for units. If we multiply each number in the set by the same constant, for example, we double them all, the standard deviation will increase by the same factor. After dividing by the mean, however, any such scaling factor cancels out and the measure is independent of units. It would not matter whether we measured incomes in dollars, thousands of dollars or pounds sterling, we would get the same coefficient of variation.

As a measure of inequality, the coefficient of variation incorporates implicit value assumptions. Since deviations of income about the mean are squared before being summed in equation (5.2), an income level that is double the distance from $\bar{y}$ effectively contributes four times the weight to the index. There is little reason why this should represent an appropriate degree of aversion to income inequality. Rather, squared deviations are often used to measure the dispersion of a distribution, since they have desirable statistical properties.

There are two other properties of the coefficient of variation that are, however, significant indicators of the ethical values that could underlie this measure of inequality. First, as shown in the final expression in equation (5.3), the coefficient of variation treats incomes that are above the mean level symmetrically with incomes that are below the mean level by the same proportion. Proportional differences between each income and the mean are then squared before being summed. Thus, an income that is 20% above the mean will contribute the same amount to the coefficient of variation as an income that is 20% below the mean. Second, two distributions with the same proportional dispersion around the mean will have the same coefficient of variation measure.

Both these properties show that the coefficient of variation is to a greater extent a measure of relative inequality than of an inability to achieve an acceptable standard of living. A compassionate concern for individuals who are too poor to afford even the minimum material essentials of life should make one very concerned about deviations of income below some minimum acceptable standard. It should not, however, make one concerned about dispersion in incomes *per se* and, in particular, about variations in incomes among people who all are well above minimum acceptable standards.

For a compassionate person, concern about high incomes is at best a derived concern that must be based on claims about what might be feasible policy actions in practice. If incomes could be redistributed without compromising other desirable objectives, the persistence of above-average incomes might indicate that the poorest individuals could be made better off through further redistribution. A compassionate person may react negatively to above-average incomes because of what such incomes might imply about unfulfilled possibilities for improving living standards of the very poor. Compassion, however, does not justify reacting negatively to above-average incomes in and of themselves. In particular, the coefficient of variation is increased by a pure Pareto improvement that benefits only those who are earning incomes that are above average. For example, suppose in the above example that the highest income is increased from 11 to 16 while the other incomes are left unchanged at 0, 2, 5 and 7. The mean income becomes (0+2+5+7+16)/5 = 6, the variance more than doubles

$$\sigma^2 = \frac{(0-6)^2 + (1-6)^2 + (5-6)^2 + (7-6)^2 + (16-6)^2}{4} = 40.75 \qquad \text{(EQ 5.5)}$$

and the coefficient of variation becomes $\sqrt{(40.75)}/6 = 1.064$ instead of 0.9. A measure of inequality that treats incomes above the mean in a way that is symmetrical to incomes below the mean, and treats reducing high incomes as a desirable goal or objective, must be based on some ethical principle other than a compassionate concern for the poor.

Similarly, by measuring inequality in terms of proportional deviations around the mean, the coefficient of variation denies the value, in terms of reducing poverty, of a uniform percentage increase in all income levels. Again a compassionate concern for the poor should imply a reduced concern about inequality as the material standard of living of everyone in the society increases.

The other widely used measure of income inequality, the Gini coefficient, can be criticised on similar grounds. There are a number of ways of viewing the Gini coefficient. The usual representation is graphical. First, the population is arranged in terms of increasing levels of income. One then calculates the percentage of total income that is earned by the bottom x% of the population for all values of x between zero and 100. These income proportions are then graphed against x to obtain the Lorenz curve. Since 0% of the population earns 0% of the total income, and 100% of the population earns 100% of the total income, the Lorenz curve runs from one corner of the unit square to the diametrically opposite corner, as illustrated in Figure 5.1.

If everyone had the same income, the Lorenz curve would correspond to the diagonal of the square: 20% of the population would earn 20% of the total income, 50% of the people 50% of the income and so on. Inequality results in a Lorenz curve that falls below the diagonal. In Figure 5.1, for example, the 50% of the population with the lowest incomes earns about 25% of total income, while the bottom 75% earns about 50% of the total income.

For the sample economy discussed above, with five people earning 0, 2, 5, 7 and 11, the Lorenz curve can be calculated as follows. One person, or 20% of the population earns 0% of the income. Two people, or 40% of the population, earn 2/25, or 8%, of the total income. Three people, that is 60%

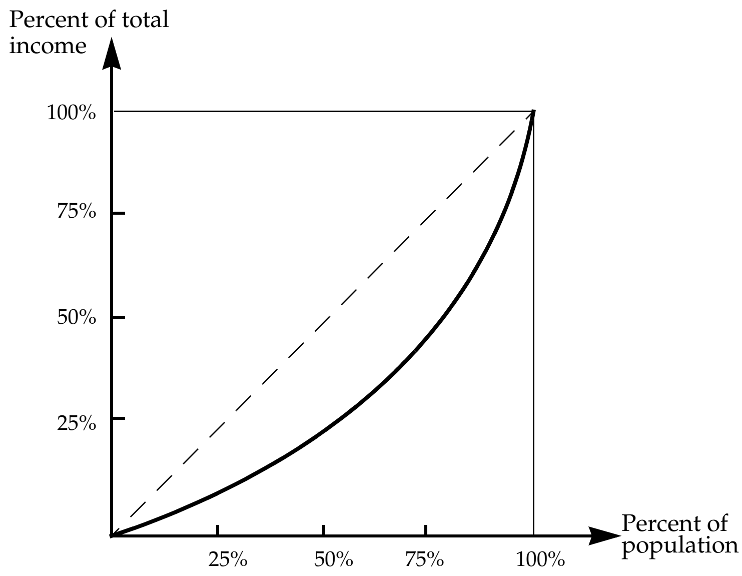

FIGURE 5.1: A Lorenz curve

of the population, earn 7/25, or 28% of the total income. Four people, or 80% of the population earn 14/25, or 56% of the total income. Finally, of course, all five people, or 100% of the population, earn 100% of the total income. The result is a Lorenz curve as in Figure 5.2.

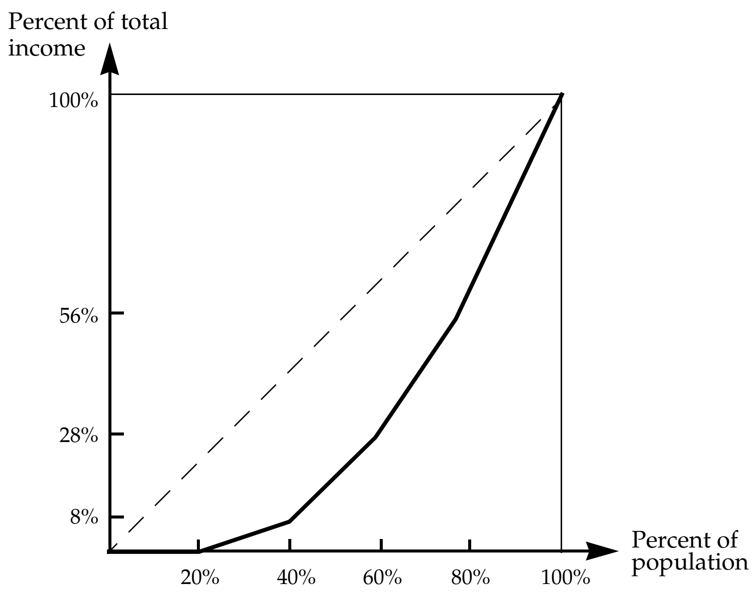

FIGURE 5.2: The Lorenz curve in the imaginary economy

Statistics New Zealand (1999: 59, Figure 4.8) presents Lorenz curves for household market income and "equivalent household disposable income" (Figures 5.2 and 5.3 on pages 68–69) in New Zealand. In the latter case, incomes are first adjusted for the effects of taxes and transfers and then adjustments are made for household size and composition to allow for economies of scale in household production.

The Gini coefficient is measured as a ratio of two areas. The numerator of the ratio is the area between the diagonal of the unit square and the Lorenz curve. The denominator is the area beneath the diagonal, that is one half.

In the sample economy, where five people earn 0, 2, 5, 7 and 11, the Lorenz curve was graphed in Figure 5.2. The area under that curve is a sum of triangles and rectangles that can be calculated to be 0.284. Hence the Gini coefficient is (0.5–0.284)/0.5 = 54/125 = 0.432.

If everyone had the same income, the Gini coefficient would be zero. If one person earned all the income, the Lorenz curve would correspond with the x-axis until $x = 100\%$, at which point it would jump to the top right corner of the unit square. In this case, the Gini coefficient would be 1.

Statistics New Zealand (1999: 27) reports that Gini coefficients for wages and salaries of people in New Zealand aged 15–64 are about 0.48. The coefficients are lower for males than for females, probably because more females are in part-time employment. The Gini coefficients are somewhat higher for total market income, and substantially so (to 0.61) if one considers all people aged 15 and above (including those without market income). Disposable income for all people aged 15 and above, however, has a Gini coefficient of about 0.46, while adjustments for household size and composition further reduce the Gini coefficient to 0.32.

It is instructive to compare the Gini coefficients in the two special cases illustrated in Figure 5.3. In case 1, we have a society in which 20% of the population has a zero income, while the remaining 80% of the population have identical incomes, making the Lorenz curve two straight line segments. In case 2, a single person gets 20% of total income while the remaining 80% of income is shared equally among virtually 100% of the population. The Lorenz curves in these two cases, graphed in Figure 5.3, show that the Gini coefficients will be identical. The areas of the two triangular gaps between the Lorenz curves and the diagonal of the unit square will be identical.

While the Gini coefficients in the two cases in Figure 5.3 are equal, the implications for concern about inequality are very different. Someone who was motivated by compassion for people with insufficient access to material resources would be most concerned about the economy in which 20% of the

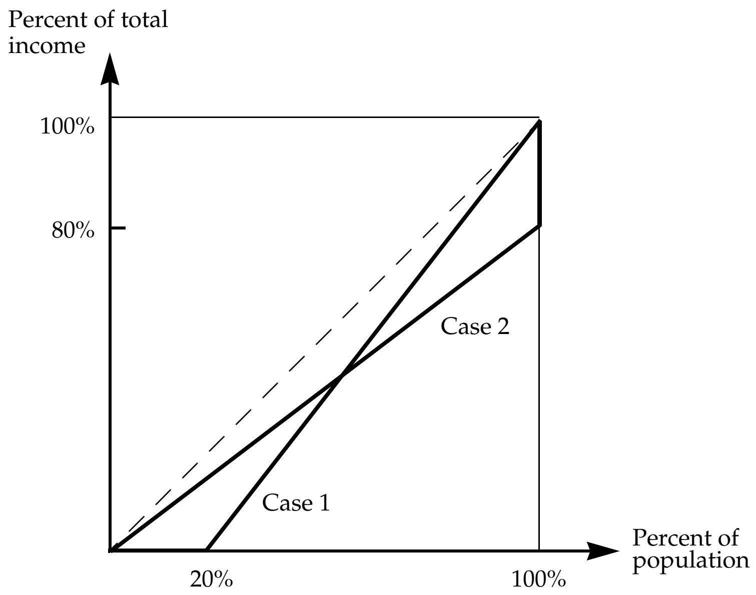

FIGURE 5.3: Lorenz curves in two special cases

population have zero income. As long as total income in case 2 were high enough to ensure that the 80% with equal incomes were not living in poverty, such a person should not be concerned about the inequality in case 2. Again, the Gini coefficient only makes sense as a measure of inequality if one accepts envy of the rich as equally worthy of concern as compassion for the poor.

The appendix to this chapter uses algebraic expressions for the Gini coefficient to illustrate that it treats deviations above the mean symmetrically with deviations below the mean. The appendix also uses the algebraic formulations of the Gini coefficient to relate it to both the coefficient of variation and to the Rawlsian maximin objective function.

As a measure of inequality, the Gini coefficient appears to reflect an envy of high incomes as much as a compassionate concern for the poor. A Pareto improving reallocation of resources that benefits only those who are earning more than the average level of income would increase the Gini coefficient. For example, in our imaginary economy with five people earning 0, 2, 5, 7 and 11, if we increase the highest income to 16 instead of 11, the Gini coefficient rises from 0.432 to 0.493. Equivalently, both the coefficient of variation and the Gini coefficient would measure a reduction in the incomes of the highest earners that provides nothing to anyone else as a desirable change. It could only be considered desirable to harm someone while doing nothing for anyone else if people are envious of those with higher incomes.

Many supporters of income redistribution emphasise their concern for the poor, but use the coefficient of variation or the Gini coefficient to measure success in achieving their goal. These measures are not appropriate, however, for gauging the extent of poverty in society. The coefficient of variation and the Gini coefficient measure the degree of inequality, not the pervasiveness of poverty.

Compassion, envy and social welfare functions

The difference between a concern for inequality and a concern for poverty can be illustrated using the approach of Atkinson (1970). Suppose we had a function $W(y_1, y_2, \ldots, y_n)$, called a social welfare function, that gives a ranking of income distributions for a population of n people. As long as function W depends only on income levels and not individual identities, it could take a wide variety of forms. The discussion of the Gini coefficient in the appendix to this chapter shows that it is one particular form of such a function, albeit one with some rather peculiar properties.

The social welfare function $W(y_1, y_2, \ldots, y_n)$ is usually assumed to be strictly increasing in all its arguments. As a result, a change that is a pure Pareto improvement, so at least one individual is made better off and no one else is made worse off, will always increase social welfare.

The social welfare function is the most widely accepted framework among economists for formalising the trade-off between equality and efficiency. Under this framework, concerns about inequality can never be strong enough to result in a negative evaluation of a pure Pareto improvement. Since the Gini coefficient can decline in response to a pure Pareto improvement, economists who advocate using the Gini coefficient as a measure of inequality usually supplement the Gini coefficient with another measure such as the average level of income. They then allow explicitly for a trade-off between equality and efficiency by weighting the Gini coefficient and the average level of income to arrive at a composite measure of social welfare.

For a given distribution $y_1, y_2, \ldots, y_n$, Atkinson defines an 'equally distributed equivalent income' as the equal income level y_e that would yield the same value for W as the actual distribution produces, so that y_e solves:

$$W(y_e, y_e, \ldots, y_e) = W(y_1, y_2, \ldots, y_n). \tag{EQ 5.6}$$

Atkinson's measure of inequality is then defined as

$$A = 1 - \frac{y_e}{\bar{y}}. \qquad \text{(EQ 5.7)}$$

The intuitive implication of this measure might be gauged by examining the special case of two people with incomes y_1 and y_2. We might suppose that the function $W(y_1, y_2)$ reflects a preference for equality but also a preference for higher total incomes. Holding one income fixed, an increase in the other income would raise W, but by progressively lesser amounts per dollar of increased income as just one of the income levels rises.

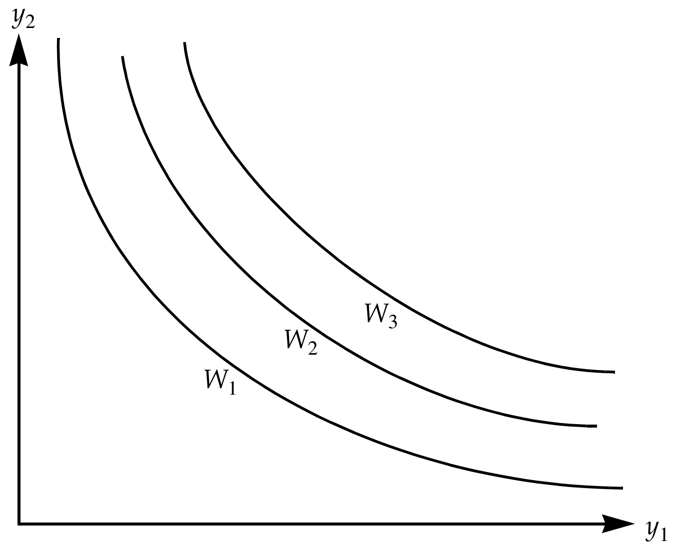

FIGURE 5.4: Income distributions of equal value

We can graph combinations of y_1 and y_2 that produce a given value of W (the equivalent of indifference curves in economics, which are combinations of consumption levels of different goods that yield a constant level of utility). For a given value of W, the result will be a curve in (y_1, y_2)-space that is 'bowed in' toward the origin as in Figure 5.4. As either y_1 or y_2 is increased we would move to curves with progressively higher levels of W. The rate of increase of W per dollar of increased income will decline, however, if only one of the incomes is increased (so bigger increases in total income are required to achieve a given increase in W). The assumption that the identities of the individuals is irrelevant will imply that the curves in Figure 5.4 are symmetrical about the 45° line.

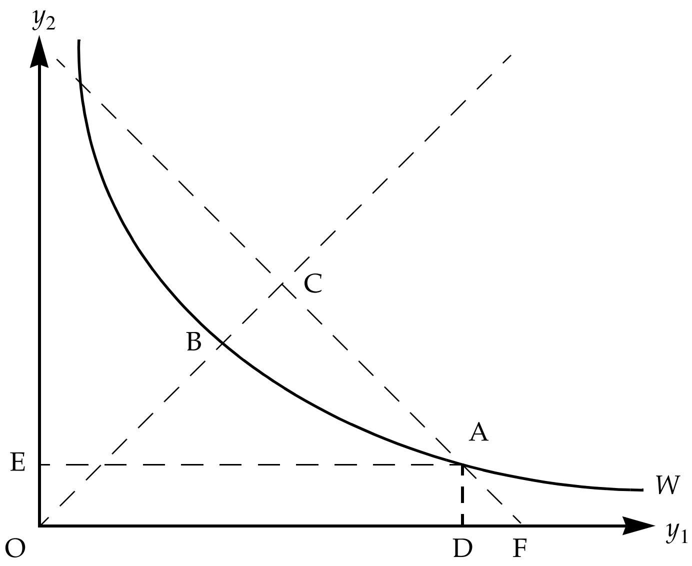

FIGURE 5.5: Illustration of Atkinson's measure of inequality

Suppose we are considering a distribution of income as at point A in Figure 5.5. The initial income of individual 1 is the distance OD, while the initial income of individual 2 is the distance OE. The line through A and C is perpendicular to the 45° line, so point C represents the average of the initial incomes. (One way of seeing this is to note that DF equals DA, so that OF equals y_1+y_2 while a vertical line from C to the y_1 axis would bisect OF.) The equally distributed equivalent income y_e will be represented by the income at the point B where the *W* curve passing through A intersects the 45° line. Incomes are equal at B and the value of *W* is the same as it is at A, as required by equation (5.6). Atkinson's measure in this case is, from equation (5.7), given by 1 – OB/OC. The more the *W* curve through A bends in toward the origin, the greater will be the inequality measure associated with distribution A.

Indifference to the distribution of income would be represented by straight line *W* curves that are perpendicular to the 45° line. Only the average income matters. Its distribution between y_1 and y_2 makes no difference to *W*.

The opposite case to a *W* function in which distribution does not matter is, in a sense, the maximin criterion of Rawls (discussed in the appendix to Chapter 3). In the maximin case, when $y_1 < y_2$ *only* an increase in y_1 can increase *W*, and conversely when $y_1 > y_2$. The curves in Figure 5.4 all become 'maximally bowed in' right angles, as illustrated in Figure 5.6.

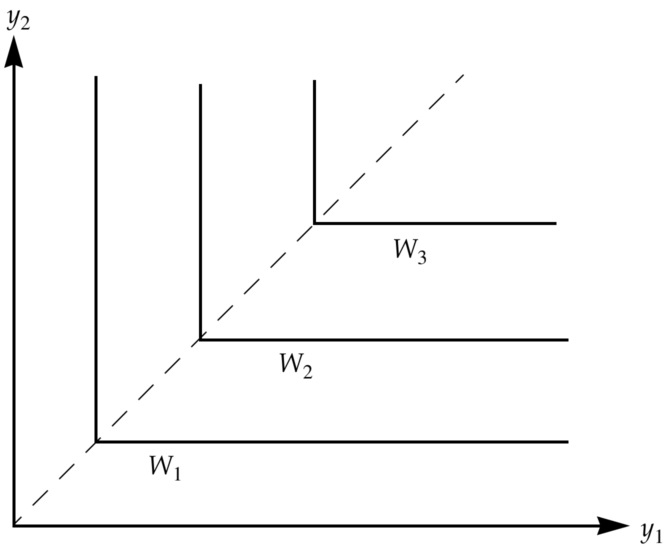

FIGURE 5.6: The maximin W function

What, in this framework, corresponds to the scale independence property of the coefficient of variation and the Gini coefficient? Specifically, we saw that in each of those cases a uniform proportional increase in all incomes would leave both indices unchanged. They were both defined in terms of relative deviations of income about the mean level $\bar{y}$. Suppose there is some coefficient $\mu > 0$ such that:

$$W(y_1, y_2, \ldots, y_n) = k^{\mu} W\left(\frac{y_1}{k}, \frac{y_2}{k}, \ldots, \frac{y_n}{k}\right) \qquad \text{(EQ 5.8)}$$

for *any* multiplier $k > 0$. Then if $W(y_1, y_2, \ldots, y_n) = W(y_e, y_e, \ldots, y_e)$

$$W\left(\frac{y_e}{\bar{y}}, \frac{y_e}{\bar{y}}, \ldots, \frac{y_e}{\bar{y}}\right) = W\left(\frac{y_1}{\bar{y}}, \frac{y_2}{\bar{y}}, \ldots, \frac{y_n}{\bar{y}}\right) \qquad \text{(EQ 5.9)}$$

and $y_e/\bar{y}$ will be the equally distributed equivalent income to the distribution $y_1/\bar{y}, y_2/\bar{y}, \ldots, y_n/\bar{y}$. Under these circumstances, any proportional increase in all incomes that leaves $y_i/\bar{y}$ unchanged for each i will also leave $y_e/\bar{y}$, and hence Atkinson's measure (equation (5.7)), unchanged.

In the two incomes case, the property (equation (5.8)), which is known as *homotheticity*, allows the W function to be written as a function of the ratio of y_1 to y_2 and then the level of mean income. Geometrically, the curves in Figure 5.4 are parallel along any given ray from the origin. The ray fixes the ratio of incomes, while the mean income level determines how far out along that ray the current income levels are. The curves corresponding to different levels of *W* are just rescaled versions of each other as illustrated in Figure 5.7.

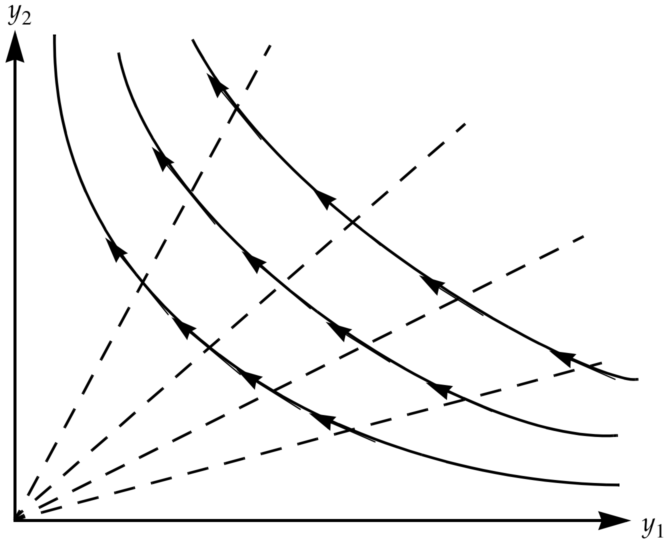

FIGURE 5.7: A homothetic *W* function

Under the homotheticity property, we have, in a sense, separate measures of the income level and inequality (see also the discussion in Creedy (1997), Chapter 2). An increase in $\bar{y}$ holding the relative incomes fixed corresponds to a move along a ray and represents an unambiguous improvement in *W*. A change in distribution that raises equality is represented by a rotation in the ray we are on toward the 45° line.

Figure 5.8 illustrates a non-homothetic *W* function. In this example, there is a minimum income level y_{min} needed to obtain an acceptable standard of living. As the income of either individual approaches that level, increases in that person's income yield large increases in *W*. As $\bar{y}$ declines toward y_{min}, and for $\bar{y} \leq y_{min}$, the *W* function indifference curves in Figure 5.8 become the maximin curves of Figure 5.6. When only one individual is below the minimum acceptable income level, raising that person's income becomes an absolute priority. If they are both below the minimum acceptable standard, there is a strong desire to share the limited available income equally. The *W* curve is either vertical or horizontal. Once both incomes move away from y_{min}, however, there is increasing indifference as to which person obtains any further income increase. Within a broad middle region, which becomes larger as average income continues to rise above y_{min}, *W* is unaffected by any change in the ratio of y_1 to y_2.

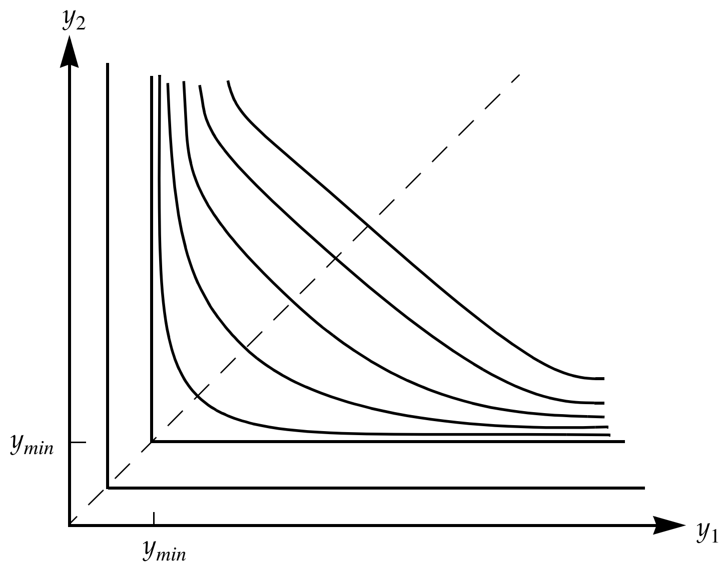

FIGURE 5.8: A non-homothetic W function

Equivalently, moving out from the origin along any fixed ray, that is for any given ratio of y_1 to y_2, eventually both incomes become large enough to put us in the middle region where the slope of the W curve is perpendicular to the 45° line. Increases in y_2 then have the same effect on W as equivalent increases in y_1. Only efficiency, or the total wealth of both people, matters and distribution becomes irrelevant.

We suggest that a W function such as the one graphed in Figure 5.8 more adequately represents the notion that compassion for people in genuine poverty motivates our concern for inequality. In such a world, once the level of all incomes gets sufficiently above the minimum acceptable level, inequality should not matter.

A numerical example might help clarify this point. Consider the ray $y_2 = 50y_1$. When y_1 = \$1,000 per annum and y_2 = \$50,000 per annum, most people would think it beneficial to redistribute from individual 2 to individual 1. (We are assuming implicitly here that other effects of redistribution can be ignored, and that persons 1 and 2 are in otherwise similar circumstances – for example, 1 is not a teenager living at home and enjoying parental support.) Thus, suppose we agree that the welfare function is very steep at y_1 = \$1,000 and y_2 = \$50,000, so that W can be raised only a little by further increasing y_2 but that even small increases in y_1 would raise W substantially. It does not follow, however, that our attitude to redistributing a dollar from person 2 to person 1 will be exactly the same when 1 is earning $500,000 per annum and 2 is earning $25 million. Yet that is what homotheticity would imply.

A W function such as the one represented in Figure 5.8 thus represents a compassionate concern for the poor, but rejects envy as a legitimate basis for redistributing income. As long as everyone has an acceptable material standard of living, we do not care much about what happens to the remaining income. Under this type of welfare function, distributional issues cannot be separated from the average level of income. There is a difference between a concern for inequality that is based on compassion and one that is based on envy.

Caring about variance and skewness

The implications of different W functions become much more difficult to appreciate when we compare more than two individuals and therefore cannot draw indifference curves as in Figure 5.4–Figure 5.8. We also need a simple way of representing attitudes to different distributions of income that can be based on a small number of easily measured statistics. This may explain the continuing popularity of measures such as the coefficient of variation or the Gini coefficient, despite their conceptual limitations.

A similar problem arises when individuals are choosing among alternatives, and they are uncertain about the consequences of each choice (see the discussion of expected utility theory in the appendix to Chapter 3). A practical example involves investing in financial assets that have uncertain returns. Economists have found it useful to focus on easily measured characteristics of returns, such as their means and variances.

We can apply the same ideas to welfare function evaluations of income distributions. Recall the discussion of expected utility theory in the appendix to Chapter 3. In Chapter 5 the distribution of income over people corresponds to the probability distribution defined on the outcomes y_i for a given person behind the veil of ignorance in Chapter 3. We can also associate the welfare function W with the utility function U in Chapter 3.

As in the appendix to Chapter 3, we can associate aversion to inequality with risk aversion. A distribution of income A could then be said to be more unequal than another distribution B with the same mean if distribution A has a higher variance than distribution B. The coefficient of variation attempts to avoid the reference to mean values in this statement by measuring variances (actually their square roots) as a proportion of the mean values.

We argued that there is a difference between a concern for inequality that is based on compassion and one that is based on envy. This implies, however, that variances, or even variances and means together, are insufficient to measure inequality. One also needs to consider the *skewness*, or third moment of the distribution about the mean. Specifically, again suppose we have n people with incomes y_1, y_2, …, y_n. Corresponding to the sample mean (equation (5.1)) and sample variance (equation (5.2)) the sample skewness of the distribution, or third moment normalised by the standard deviation, is defined by:

$$s^3 = \frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}\left[\frac{(y_i-\bar{y})}{\sigma}\right]^3. \qquad \text{(EQ 5.10)}$$

As argued in the appendix to Chapter 3, skewness becomes important when there is a minimally acceptable material standard of living. In particular, it then becomes extremely important to worry about possible situations in which income may approach the minimally acceptable standard.

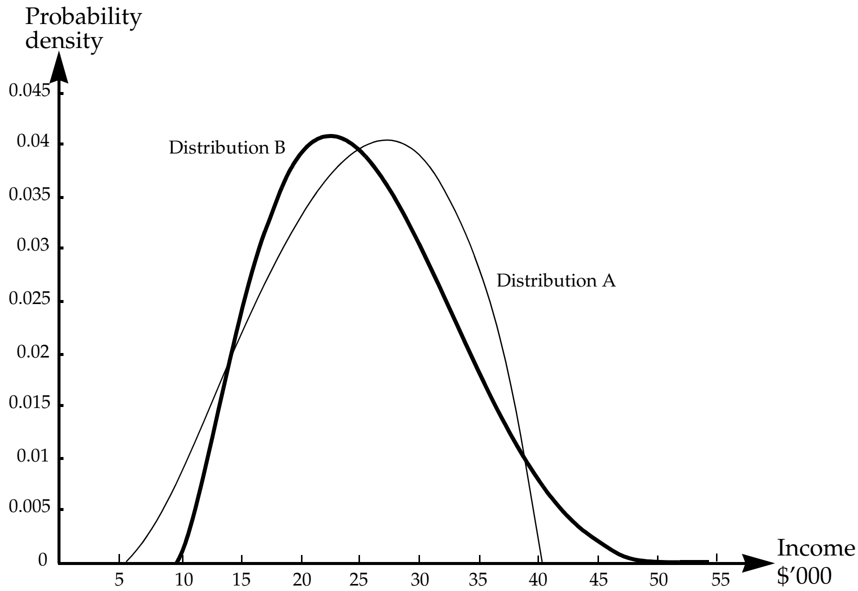

FIGURE 5.9: Income distributions with the same mean, coefficient of variation and Gini coefficient

The role of skewness might be clarified with the aid of an example. Consider the two income distributions represented in Figure 5.9. The numbers along the horizontal axis represent annual income in thousands of dollars. The functions graphed in the figure are the probability density functions. Loosely speaking these functions can be thought of as the proportion of people who earn each income level. (Technically speaking, the proportion of people earning income within a given band or range is the integral of the density function over that range.)

The income distributions in Figure 5.9 have been constructed to have the same annual mean income of $25,000, the same coefficient of variation of 0.30 and the same Gini coefficient of 0.08. They would thus be regarded as 'equivalently unequal' using any of the usual measures of inequality. While the minimum income under distribution B is $9,261, compared with only $5,020 under distribution A, the maximum income under distribution B is $53,936, compared with only $40,214 under distribution A. Furthermore, the modal (or 'most likely') income (where the density function peaks) is about $5,000 higher under distribution A, which also has a broader 'middle income' group.

In our view, however, the two distributions could be considered equivalently unequal only if envy is just as legitimate a motivation for redistribution as is compassion. One gets from distribution B to distribution A by eliminating some high incomes, while adding some very low incomes. There is also a tendency for the middle income group to be relatively better off under distribution A. Yet if one's concern about income inequality is motivated primarily by a concern for the very poor, the increase in low incomes in moving from B to A should be far more troubling than any gain from making the reasonably well off under distribution B a little better off in relative terms under distribution A.

The key difference between distributions A and B is that they have different signs of skewness. Instead of being symmetrical about the mean, incomes under distribution A are skewed toward values that are low relative to the mean. Conversely, distribution B has incomes skewed toward values that are high relative to the mean. In fact, the skewness of distribution A is –0.19844 compared with +0.38689 for distribution B. Any concern for inequality that draws a distinction between envy and compassion as motivating factors would take note of the skewness of the distribution. Extremely low incomes are the primary focus of a compassionate person, while an envious person would focus on the extremely high outliers.

Redistribution should be motivated by compassion for our fellow human beings who find themselves in distress, rather than a desire to achieve equality:

> There has long been a widely supported belief that we all have a duty to prevent starvation; in Britain from medieval times until the twentieth century this obligation was placed upon local public authorities. The problem today is that the wide support for the relief of hardship has been exploited by egalitarians to cultivate support for equality, a doctrine calling for the power of the state to be used not to assist the poor, but to equalise people.
>
> Advocates of equalisation have consequently been able to exploit support for the relief of poverty by falsely characterising measures designed to equalise people as measures to relieve suffering. (Green, 1996: 46)

Why not indulge envy?

We have emphasised that much support for income redistribution appears to be based on envy even though the rhetoric used to support those policies appeals to compassion. It is necessary, therefore, to discuss why it is wrong to base income redistribution policy on envy.

As we noted earlier, income and wealth are at best very imperfect measures of welfare:

> Two sisters born in Nebraska travel to New York to find careers. As it happens, not without heartbreaks along the way, neither sister marries, but both find jobs. Alice, the older one, retains the thrifty habits of her Nebraska childhood, lives nicely, and steadily puts aside significant sums every year. She invests in a growth fund (one-third) and a blue chip mutual fund (two-thirds), considering herself a little more modern and adventuresome than her father, who went through World War I and the Great Depression.
>
> Meanwhile, her sister Sarah thinks Alice is crazy. "You don't have any children. What are you saving for? You should live a little," she tells Alice often – too often. Sarah spends everything she has. She loves eating out, wearing fine clothes, furnishing her apartment expensively. Her passion is travel. She has been to the Himalayas, to Constantinople, to the Upper Congo, to India, to Australia. The only thing she saves for is travel – an account she then drains at least every other year for another fabulous trip. She likes going alone, her own way. She knows what she enjoys and plans for a certain degree of luxury, allowing time and space for surprise and adventure.
>
> By the time they reach their sixties, Alice, it turns out, has over a million dollars in her mutual funds (not counting her generous pension from work). Sarah has no more than $5,000 in all her accounts put together, including her kitchen piggy bank. Do these sisters envy each other? On the contrary. They feel sorry for one another. Each thinks the other is crazy. (Novak, 1996: 91–92)

People can differ in many dimensions. There is no reason to give special recognition to feelings of envy about income or wealth when we would not countenance appeasing feelings of envy relating to other personal characteristics. Most people would baulk at the notion, for example, that envy of the blind for the sighted ought to be rectified by destroying, or even handicapping, the vision of those who can see.

There is one difference between envy of the rich and envy of the sighted. The loss in overall social welfare is obvious in the case of blinding a sighted person, whereas many proponents of equality often assume that money can be transferred from one person to another at no cost. A dollar taken from Jones is still a dollar when it is given to Smith. This ignores the costs of administering the tax and welfare systems and the substantial efficiency losses accompanying both taxes and transfers. Since a tax discourages Jones from doing something with benefits in excess of its cost, the efficiency losses from a dollar of tax revenue are more nearly $1.25 than a dollar in New Zealand as we noted in Chapter 4. Similarly, a means-tested transfer encourages Smith to substitute untaxed activities for market labour. The marginal cost to Smith of supplying market labour will be less than the marginal social benefit of Smith doing so. Smith will also be encouraged to adjust affairs in other ways, such as a change in marital status or place of abode, that may increase Smith's eligibility for the transfer payments and yet reduce Smith's utility. The fraction of a dollar that Smith receives from an extra dollar of taxes (after subtracting administrative costs) is actually worth less to Smith than the monetary value of the transfer.

Income can, of course, partially compensate for other disadvantages. In principle, one could adjust income measures for disadvantages such as blindness and base redistribution policies on the corrected measures. This is, however, impossible to do in practice. There is no way of knowing what the appropriate correction ought to be. For example, the appropriate compensation for blindness might depend upon how one is affected psychologically by this disability. These psychological states would be impossible to convey to anyone else with any credibility.

We also noted earlier in this chapter that many differences in income are chosen by people because they prefer non-monetary rewards to higher income. Redistribution of income at the middle-to-upper ends of the distribution that is aimed at reducing inequality measures such as the Gini coefficient or the coefficient of variation may exacerbate rather than reduce inequalities in standards of living.

Even an egalitarian such as Temkin (1993: 17) acknowledges:

> I think deserved inequalities are not bad at all. Rather, what is objectionable is some being worse off than others through no fault of their own. Thus, although there may be more inequality in one situation than in another, that needn't be worse if the lesser inequality is deserved but the greater is not. Correspondingly,

for each of this book's examples I shall assume that people are equally skilled, hard-working, morally worthy and so forth, so that those who are worse off than others are so through no fault of their own.

Temkin goes on to remark:

> It is, of course, extremely difficult to decide when people are worse off than others through no fault of their own. Some think this is nearly always the case, others, almost never the case. Fortunately, one need not decide this issue (which involves, among other things, the mare's nest of free will) in order to recognize that only undeserved inequalities are bad.

One does need to decide the issue, however, in order to implement any policies aimed at the goal of reducing inequality. Measures of observed income reflect not only chance events or innate abilities but also myriad choices that people make. We typically need long and expensive trials involving a prosecution, defence, jury, judge, and established rules of evidence in order to decide intentionality in single acts of alleged criminal behaviour. It is unrealistic to suppose that one could separate observed annual income measures into one component reflecting undeserved innate ability and another reflecting deserved consequences of countless decisions made every day. Furthermore, any attempt to do so would impose costs that are unlikely to be justified by the benefits of appeased envy, whatever they may be.

In this regard, there is a significant difference between redistribution based on compassion and redistribution based on envy. The type of information that one needs in order to implement redistribution based on compassion is very different from the information needed to erase undeserved differences in welfare. In the former case, we recognise that a minimal level of money income, deserved or not, is necessary in a modern market-oriented economy for someone to have a standard of living beneath which no one should fall. As we discuss later in this chapter, there are practical difficulties with determining such a level of income. These problems are minimal, however, compared with the practical difficulties of redistributing money incomes so as to reduce inequality of welfare. As Jouvenal (1952: 30) noted:

> Redistribution started with a feeling that some have too little and some too much. When attempts are made to express this feeling more precisely, two formulae are spontaneously offered. The first we may call objective, the second subjective. The objective formula is based upon an idea of a decent way of life beneath which no one should fall and above which other ways of life are desirable and acceptable within a certain range. The subjective formula is not based upon a notion of what is objectively good for men but can be roughly stated as follows: 'The richer

> would feel their loss less than the poorer would appreciate their gain'; or even more roughly: 'A certain loss of income would mean less to the richer than the consequent gain would mean to the poorer.'

Suppose, however, that we put aside the practical difficulties of adjusting income measures to allow for merit and inequalities in other dimensions. There are still fundamental objections to the notion that feelings of envy ought to be appeased.

Nozick (1974: 239–246) traces envy to another emotion, namely self-esteem. He defines envy as follows:

> The envious person, if he cannot (also) possess a thing (talent, and so on) that someone else has, prefers that the other person not have it either. The envious man prefers neither one having it, to the other's having it and his not having it. (Nozick, 1974: 239)

Nozick then asks the question:

> Why do some people *prefer* that others not have their better score on some dimension, rather than being pleased at another's being well-off or having good fortune; why don't they at least just shrug it off? One line [that] seems especially worth pursuing … [is] when the other person's having a higher score than himself threatens or undermines his own self-esteem and makes him feel inferior to the other in some important way … [W]e evaluate how *well* we do something by comparing our performance to others. (Nozick, 1974: 240, emphasis in original)

Nozick points out that this reaction may not depend on whether the envious person thinks the superior position is deserved. On the one hand:

> That there is someone else who is a good dancer will affect your estimate of how good you yourself are at dancing, even if you think that a large part of grace in dancing depends on unearned natural assets. (Nozick, 1974: 241)

On the other hand, if the superior position is deserved and earned:

> It may injure one's self-esteem and make one feel less worthy as a person to know of someone else who has accomplished more or risen higher. (Nozick, 1974: 241)

According to Nozick, a major reason for rejecting envy-based policies is that such policies are unlikely to succeed in their objective, and may even be counterproductive when measured in their own terms:

> If, after downgrading or equalizing one dimension, say wealth, the society comes generally to agree that some *other* dimension is most important … then the phenomenon will repeat itself.
>
> People generally judge themselves by how they fall along the most important dimensions in which they *differ* from others. People do not gain self-esteem from their common human capacities by comparing themselves to animals who lack them. ("I'm pretty good; I have an opposable thumb and can speak some

> language.") … When everyone, or almost everyone, has some thing or attribute, it does not function as a basis for self-esteem. Self-esteem is based on *differentiating* characteristics; that's why it's *self*-esteem. (Nozick, 1974: 243, emphasis in the original)

Thus, one is unlikely to appease envy by giving in to its demands in any one particular domain. Worse than that, however, Nozick suggests that one might worry:

> … *if* the number of dimensions [along which people can vary and that people hold to be valuable] is not unlimited and if great strides are made to eliminate differences, that as the number of differentiating dimensions shrinks, envy will become more severe. For with a small number of differentiating dimensions, many people will find that they don't do well on *any* of them. (Nozick, 1974: 245, emphasis in original)

By contrast, when there are a number of independently distributed valued attributes (either innate or acquired):

> Everyone might view themselves as at the upper end of a distribution … since each sees the distribution through the perspective of the particular weights he assigns. The fewer the dimensions, the less the opportunity for an individual successfully to use as a basis for self-esteem a non-uniform weighting strategy that gives greater weight to a dimension he scores highly in. (Nozick, 1974: 245)

Furthermore:

> The most promising ways for a society to avoid widespread differences in self-esteem would be to have no common weighting of dimensions; instead it would have a diversity of different lists of dimensions and of weightings … Such a fragmentation of a common social weighting is not to be achieved by some centralized effort to remove certain dimensions as important. The more central and widely supported the effort, the more contributions to it will come to the fore as the commonly agreed upon dimension on which will be based people's self-esteem. (Nozick, 1974: 245–46)

In a society based on free markets, people can organise themselves into many groupings, including clubs and voluntary organisations, and feel esteem due to their high rankings in their chosen subgroups. With enough valued attributes, and enough different weightings of those attributes, most people will find some weighting that ranks them high enough to give them self-esteem.

A useful analogy can be drawn between equity and trade. By doing what we do best, and trading for the other things we desire, we gain access to more opportunities than we would have by trying to be self-sufficient. The gains from trade, or from cooperating, will be greatest when people have different bundles of attributes. We can all specialise in doing the jobs for which we are relatively best suited. Reducing inequalities of attributes thus is likely to reduce our own standard of living. Should we be happy if Kiri Te Kanawa's voice is damaged because it will make us more equal, even though it means it will be less enjoyable to go to the opera?

Drawing an analogy between the generation of self-esteem and trade among countries suggests another point. *Comparative*, rather than *absolute*, advantage determines patterns of trade in a market economy. Countries in which all industries are less productive can nevertheless participate in the world economy by concentrating on those industries for which their disadvantages are relatively less significant.

Similarly, people's relative abilities will determine the value of the contribution that each person can make to society. We suggest, therefore, that relative endowments may also determine the amount of self-esteem one obtains from one's abilities.

We introduce one further assumption, namely that valued attributes are linked to productive activities either as inputs or outputs. For example, attributes such as musical skills, sporting ability, intellectual capacity or carpentry skills may be valued because they are important inputs into the production of other things that people value. Note that productive activities need not be restricted to market activities. Alternatively, the attributes that any person possesses may be enhanced by productive activities (learning, practice, grooming and so on). When the attributes are outputs rather than inputs, the best skills to develop will also depend on relative rather than absolute ability endowments. For concreteness, we shall assume that attributes are valued because they can be used to produce goods or services that people wish to consume.

Suppose we consider a society of two people each of whom has different levels of two valued attributes. Let the people be *A* and *B*, and label the attributes 1 and 2. Suppose person *A* is endowed with higher levels of both attributes 1 and 2 than is person *B*. Further suppose, however, that the endowed level of attribute 1 for person *A* is twice the endowed level of attribute 1 for person *B*, whereas person *A* has only 50% more than *B* of attribute 2. Define units so that person *B* has equal amounts of attributes 1 and 2. Person *A* will then have $^4/_3$ times as much of 1 as *A* has of 2.

Any time that person *A* spends exploiting attribute 2 could instead be spent exploiting attribute 1. Since *A* has relatively fewer advantages using attribute 2 than using attribute 1, the output *A* would get would be greater if *A* instead focused on activities that exploited attribute 1 and then traded what *A* produced with what *B* produces. Person *A* is better off concentrating on what *A* is able to do best and letting *B* concentrate on what *B* is able to do best. Person *B* can become the specialist in using attribute 2 to produce output, and get self-esteem from being that specialist, even though *B*'s absolute productivity in those activities would be less than A's productivity, were *A* to do those things instead.

A similar argument would apply if people could allocate time and other resources to developing different attributes or skills. In general, it will pay each person to develop those skills that are possessed in relatively greater abundance. Even someone who could be more productive at developing all types of attributes may be better off concentrating effort on developing the one attribute most likely to yield the greatest beneficial outcome.

Finally, perhaps the strongest argument against basing policy on envy, supposing that it could be done successfully, is that such a stance would be a serious infringement of individual freedom for a dubious social gain. As Nozick argues:

> Even if envy is more tractable than our considerations imply, it would be objectionable to intervene to reduce someone's situation in order to lessen the envy and unhappiness others feel in knowing of his situation. Such a policy is comparable to one that forbids some act (for example, racially mixed couples walking holding hands) because the mere knowledge that it is being done makes others unhappy. (Nozick, 1974: 245)

The argument is related to our discussion in Chapter 2 of the potential conflict between efficiency and individual liberty when the utility of one person depends directly on the choices made by someone else. As in that case, we would argue that liberty ought to take precedence over efficiency.

ALLEVIATING POVERTY AS A GOAL

By devising measures of distributional outcomes such as poverty lines, governments have recognised that alleviating poverty is not the same as promoting equality. Poverty lines ostensibly measure the income needed to attain "a minimally adequate standard of living" (Citro and Michael, 1995: 19).

The first attempt to devise a 'poverty line' to be used for informing government welfare and taxation policies was made by a staff economist at the United States Social Security Administration (SSA) in the early 1960s. The measure "had a set of poverty thresholds for different types of families that consisted of the cost of a minimum adequate diet multiplied by three to allow for other expenses" (Citro and Michael, 1995: 17).

Citro and Michael summarise the subsequent history of the poverty index in the United States:

> In 1965 the Office of Economic Opportunity (OEO) adopted the SSA thresholds for statistical and program planning purposes; in 1969 the U.S. Bureau of the Budget (now the U.S. Office of Management and Budget) issued a statistical policy directive that gave the thresholds official status throughout the federal government. The Census Bureau took over the job of publishing the official annual statistics on the number and proportion of poor (the poverty rate) by comparing the SSA thresholds to estimates of families' before-tax money income from the March Current Population Survey (it first issued poverty statistics in August 1967). For these comparisons, the SSA thresholds are updated annually for price inflation and so are not changed in real dollar terms. (Citro and Michael, 1995: 17)

The Panel on Poverty and Family Assistance, which produced the book edited by Citro and Michael, was funded by the United States Census Bureau to examine the adequacy of the then-current poverty measure for the United States. They concluded that a new measure was required:

> Our measure includes a specific concept of economic poverty by which to develop a new poverty threshold for a reference family type: inadequate resources to obtain basic living needs. We define those basic needs as food, clothing, and shelter. There are other needs as well (e.g., personal care, transportation), but there is less agreement about them, and so our approach provides a small amount for other needed spending by means of a multiplier that is applied to the amounts for food, clothing, and shelter.
>
> This concept of poverty as insufficient resources for basic living needs accords with traditional public concerns for the needy …
>
> Under our threshold concept, we propose that the values for food, shelter, and clothing – the basic bundle – and for a small amount of other needed spending – the multiplier – be developed by direct reference to spending patterns of American families below the median expenditure level. More important, we propose that real changes in spending on food, clothing, and shelter be used to update the poverty thresholds each year. By so doing, the thresholds will maintain a relationship to real changes in living standards, but only to the extent that these changes affect consumption of basic goods and services that pertain to a concept of poverty, not all goods and services. (Citro and Michael, 1995: 22–23)

The original poverty index in the United States included allowances for different types of family composition. The Panel on Poverty and Family Assistance suggested that allowance should also be made for geographic variation in costs of living and for differences in time available for the direct provision of household services.

Relative versus absolute measures of poverty

Citro and Michael (1995: 31) note that the US poverty index is an absolute standard, albeit one that is periodically replaced by "a new standard that is higher in real terms". They comment that such adjustments might be justified on the grounds that new items of expenditure may be deemed essential to attaining a minimally acceptable standard of living as general community affluence increases. They give the example of the telephone and comment that, these days:

> [O]ne could argue that having a telephone is essential in a developed country for everything from job seeking to having relationships with family and friends. (Citro and Michael, 1995: 22)

Changes in general levels of affluence may affect how much income is needed to function adequately in the society and thus to avoid material hardship. Changes to the officially measured poverty line, however, ought to be explicitly justified on this basis. A relative definition of poverty that fully and routinely adjusts the poverty line for changes in average income levels does not make sense. As we emphasised above, if everyone has an acceptable material standard of living, the distribution of the remaining income is irrelevant. Efficiency becomes a much more important goal than any particular distribution of monetary income.

Statistics New Zealand (1999: 79) defines "low income households" as those in the bottom 20% of "equivalent disposable" income (that is, household income after adjusting for taxes, transfers and household composition). Statistics New Zealand (1999: 87) observes, however, that:

> Conceptually, poverty can be thought of in terms of absolute poverty or in terms of relative poverty. Absolute poverty means a lack of resources sufficient to provide household members with 'adequate' nutrition, housing, and other 'essential' needs (the quotation marks signal a seemingly unavoidable fuzziness in definition). Relative poverty means that a household lacks sufficient resources to be able to fully take part in activities which most members of the community would regard as normal. Thus a household which cannot afford the fees for its children to go on school trips might not be considered to be suffering from absolute poverty, but would be poor in relative terms. In New Zealand the term tends to presume a composite of both absolute and relative poverty notions.

In practice, Statistics New Zealand (1999: 88–89) used various fractions of median equivalent disposable income as a poverty measure while noting some of its defects. Statistics New Zealand also recommended that "a poverty line should not be an official statistic" because "a poverty line tells us nothing about the degree of deprivation by those below it and risks oversimplifying our understanding of poverty". Unfortunately, the even worse defects of the Gini coefficient did not prevent Statistics New Zealand from using it uncritically as their measure of income inequality throughout their report.

In Australia and most Western European nations, poverty is officially measured using a relative standard:

> A typical approach is to select a cutoff point in the distribution of total family expenditures or income adjusted for family composition – say, one-half the median – and designate that dollar amount as the poverty threshold for a reference family, with thresholds for other family types developed by use of an equivalence scale. (Citro and Michael, 1995: 31)

Such a practice implies that poverty can be reduced only by raising the incomes of the lowest earning families at a faster rate than those initially earning higher incomes. Commenting on the situation in the United Kingdom, Green (1996: 46) observed:

> [A]ny definition of poverty must be relative to the extent that public conceptions of hardship change over time. But the fact that conceptions of hardship have changed as prosperity has increased does not justify linking the poverty line to average or median earnings. This device has the result of defining a large proportion of the population as always poor, without regard to the actual standard of living enjoyed.

Citro and Michael (1995: 31–32) put the point slightly differently:

> One criticism of relative thresholds is that the choice of the expenditure or income cutoff is arbitrary or subjective, rather than reflecting an objective standard of economic deprivation. It is also argued that relative poverty thresholds do not provide a stable target against which to measure the effects of government programs because they change each year in response to increases or decreases in real consumption levels instead of remaining fixed in real terms.

Economic deprivation is not, however, something that can be determined objectively:

> [T]he original U.S. poverty thresholds were based on … USDA estimates of the cost of the Economy Food Plan with a multiplier to account for other consumption items.
>
> Relativity and subjectivity entered into the determination of both the food component and the multiplier for the original poverty thresholds. The Economy Food Plan was developed by considering the food-buying patterns of lower income families, as well as nutritional requirements. The USDA experts could have developed the Economy Food Plan at an even lower cost level and still provided for nutritional balance if they had been willing to ignore the eating patterns of Americans, who, even at lower income levels, showed a preference for

> meat as well as rice and beans. They could also have developed the Economy Food Plan at a higher cost level to allow for somewhat greater variety of diet and an occasional restaurant meal. That is, they had to make judgments that cannot be supported by nutritional science alone; they were guided in these judgments by data on Americans' actual food choices … [the] multiplier [also] … was based on the spending patterns of the average American family rather than on expert standards for other needed budget items. (Citro and Michael, 1995: 32)

Nevertheless, an absolute poverty threshold has a number of advantages. Since the procedure is tied to explicitly named commodities and to measurable expenditure patterns, the debate about the appropriate level of the index is much more focused. The procedure also allows for the possibility that increases in average living standards can raise people out of poverty without requiring their relative incomes to increase.

The new measure of poverty in the United States that was developed by the Panel on Poverty and Family Assistance has the following characteristics (Citro and Michael, 1995: 40):

- The poverty thresholds should represent a budget for food, clothing, shelter (including utilities), and a small additional amount to allow for other needs (e.g., household supplies, personal care, non-work-related transportation).
- A threshold for a reference family type should be developed using actual consumer expenditure data and updated annually to reflect changes in expenditures on food, clothing, and shelter over the previous three years.
- The reference family threshold should be adjusted to reflect the needs of different family types and to reflect geographic differences in housing costs.
- Family resources should be defined – consistent with the threshold concept – as the sum of money income from all sources together with the value of near-money benefits (e.g., food stamps) that are available to buy goods and services in the budget, minus expenses that cannot be used to buy these goods and services. Such expenses include income and payroll taxes, child care and other work-related expenses, child support payments to another household, and out-of-pocket medical care costs, including health insurance premiums.

The proposal to update the measure for real changes in the consumption of food, clothing and shelter introduces a relative element into the measure. The panel comments, however, that the proposed measure "is not the same as a fully relative measure, such as one-half median income or expenditures, that would update the thresholds for changes in total consumption, including luxuries as well as basic goods and services" (Citro and Michael, 1995: 42).

Nevertheless, the measure currently used in the United States would appear to be too generous in adjusting for ostensible changes in the cost of meeting a minimally acceptable standard of living. For example, Table 5.1 from Cox and Alm (1999: 15) shows that families officially classified as poor

TABLE 5.1: Percent of households officially classified as poor in the United States possessing selected items

Percent of households with:	*Poor households in 1984*	*Poor households in 1994*
Washing machine	58.2	71.7
Clothes dryer	35.6	50.2
Dishwasher	13.6	19.6
Refrigerator	95.8	97.9
Freezer	29.2	28.6
Stove	95.2	97.7
Microwave	12.5	60.0
Colour television	70.3	92.5
Videocassette recorder	3.4	59.7
Personal computer	2.9	7.4
Telephone	71.0	76.7
Air-conditioner	42.5	49.6
One or more cars	64.1	71.8

in the United States in 1994 tended to own more of most consumer durables than families classified similarly in 1984, while a surprisingly large proportion of families in both years possessed a wide range of durable goods. If the current method of adjusting the poverty level is too generous, any further move towards a more relative measure may only exacerbate the problem.

New Zealand would, of course, need to develop its own index. We concur with the general approach of focusing on consumption items deemed to be necessities, but allowing for some minor increases in these expenditures as the costs of living in different regions change over time. Simply indexing a level for movements in average income or wages would not, however, be an acceptable procedure for measuring poverty.

Measuring expenditures or incomes

Green (1996: 48–49) also criticises the poverty line used in the United Kingdom on the grounds that it is based on income rather than expenditures. The Panel on Poverty and Family Assistance considered using expenditure rather than income as the basis for determining deprivation, but ultimately decided in favour of a definition based on income:

> Conceptually, an income definition is more appropriate to the view that what matters is a family's ability to attain a living standard above the poverty level by means of its own resources. Thus, an income definition will not count as poor anyone who had an income above the threshold for the period of measurement, even if he or she consumed less than the poverty level, for whatever reason – pure choice or perhaps because of anticipating a drop in future income. Conversely, an income definition will count as poor anyone who had inadequate income, even if he or she was able to maintain consumption above the poverty level by such actions as borrowing, carrying a credit card balance, or depleting savings. In contrast to an income definition, an expenditure (or consumption) definition is more appropriate to the view that what matters is someone's actual standard of living, regardless of how it is attained. In practice, the availability of high-quality data is often a prime determinant of whether an income- or expenditure-based family resource definition is used. (Citro and Michael, 1995: 36–37)

Green cites an Institute for Fiscal Studies (IFS) report (Goodman and Webb, 1995) showing that expenditure levels produce very different results for changes in the welfare of the poor than do income levels. Green comments:

> The IFS examined the possibility that the difference was the result of increased debt, but found this not to be the case. The increase in self-employment explains some of the difference … Part of the remaining difference is explained by the increase in the number of people reporting zero or negative income … In many years, the average expenditure of those reporting no income exceeded the average for the whole population, because individuals resorted to their savings.
>
> However, the chief explanation given is that the households in the bottom decile according to income vary considerably from year to year and that, in any one year, the households in the bottom tenth may have savings on which they can draw. The IFS found that, over the period 1979–1992, the proportion of pensioners in the bottom income decile fell, as they were replaced by working-age families. Some people, those unemployed for a relatively short time, for instance, draw on their savings, whereas pensioners on average spend less than non-pensioners because they are more inclined to save from their income than to spend their savings. (Green, 1996: 48–49)

There is little doubt that incomes and expenditures can differ substantially in the short run. Green (1996: 49) notes:

> A table in the 1995 edition of *Households Below Average Income* shows just how much variation there is between income and expenditure. Three percent of those in the bottom decile according to income were in the highest decile for expenditure and 20 percent were in the top 50 percent. Only 33 percent of the lowest decile according to income were in the lowest decile according to expenditure.

Statistics New Zealand (1999: Appendix A5) discusses New Zealand data on income and expenditure. Figure A5.1 in that publication shows large excesses of average expenditure over average income for the low-income households. There are also more modest excesses of expenditure over income for all but the 30% of households with the highest disposable incomes in all years examined by the report except 1982.

Similar results can be found in the United States *Consumer Expenditure Survey*. Table 5.2 presents two sets of results from the 1995–96 Survey. Money income before taxes includes public assistance, supplemental security income, food stamps, unemployment benefits and "regular contributions for support" but excludes "gifts of goods and services".

The accumulation and reduction of savings may be a significant part of the explanation for these differences between expenditure and income (see the discussion of self-insurance in Chapter 4 and the discussion of the permanent income hypothesis earlier in this chapter). Households with incomes temporarily above their long-term averages spend less than their incomes and increase their savings. Conversely, households with incomes temporarily below their long-term averages will be overrepresented in the low-income categories. Many of these households can maintain their expenditure levels by drawing down their savings or increasing their borrowing on credit cards or other types of consumer loans.

In the case of the US statistics presented in Table 5.2, this interpretation is given some credibility by the fact that, for both types of households, average self-employment income for the under \$5,000 income category is –\$2,209 for the two-person consumer units and –\$1,694 for the three-person consumer units. (These may reflect losses on small businesses or financial investments.) Average self-employment income is positive in all other income categories. Furthermore, for three-person consumer units the average net change in total assets is negative only for the \$5,000–\$9,999 income category. While the under –\$5,000 category has an increase in net assets, it also has a more-than-offsetting increase in net liabilities. Among two-person consuming units, the pattern is somewhat different. The average change in net assets and liabilities increases or decreases slightly at all income levels. While the increase in net liabilities exceeds the increase in net assets for the under –\$5,000 income bracket, the excess increase in net liabilities is much greater at the highest income levels. This may reflect an age difference in the two types of consumer units. Two-person units are more likely to be young couples taking on liabilities as they establish a new home.

TABLE 5.2: Sample US income and expenditure, 1995–1996[a]

	Consumer units of two persons		
Income before taxes	*Average before tax income $*	*Average after tax income $*	*Average annual expenditure $*
< $5,000	715	179	19,664
$5,000–9,999	7,519	7,214	17,176
$10,000–14,999	12,555	12,206	20,124
$15,000–19,999	17,369	16,887	24,584
$20,000–29,999	24,541	23,374	27,825
$30,000–39,999	34,692	32,501	33,864
$40,000–49,999	44,460	41,235	40,114
$50,000–69,999	58,657	52,992	47,296
> $70,000	106,709	94,857	70,774
Total two-person units reporting	40,337	36,916	36,581
	Consumer units of three persons		
< $5,000	1,880	1,985	21,999
$5,000–9,999	7,716	7,893	18,632
$10,000–14,999	12,357	12,356	22,654
$15,000–19,999	17,302	17,049	24,597
$20,000–29,999	24,778	23,603	29,078
$30,000–39,999	34,476	32,117	34,282
$40,000–49,999	44,920	41,614	39,423
$50,000–69,999	58,876	54,519	49,226
> $70,000	102,547	91,541	70,134
Total three-person units reporting	45,048	41,401	40,262

[a] Source: US Bureau of Labor Statistics, http://stats.bls.gov/csxhome.htm

Support from extended family members, friends and private charities may also help households maintain their expenditures in the face of lower incomes. Such support is not included in the definition of income. "Other money receipts" does, however, appear as a separate item in the original tables. While the average value is not large, it might be reasonable to suspect that, as an irregular item, cash receipts from family members may not have been accurately recorded.

In fact, errors in recording income may largely explain the fact that many households with a low recorded income nevertheless record total expenditure that is not greatly different from the average for their household type. Perhaps self-employment income, wages, and hours worked are reported with error in the survey but would be recorded more accurately when assessing tax liabilities or eligibility for transfer payments. Households recorded as receiving a low income are likely, however, to have substantial receipts from cash transactions (tips, house cleaning payments, house repairs, gardening, car repair, baby-sitting fees, possible criminal income and so forth) that are not reported to the taxation authorities, welfare agencies or to the interviewer for the survey:

> Sociologists … have been puzzled by how the poor could live on so little income. Those who have gained the confidence of the poor discover the answer: the underground economy. (Feldstein, 1999: 37)

The wide disparities between income and expenditure in any one year should make one wary of using income as the only, or even a preferred, measure of poverty. After all, if people can avoid reporting cash income to the Inland Revenue Department, or cash income or significant cash transfers to a survey interviewer, they could also avoid reporting such income to someone evaluating them for public welfare assistance.

As we noted earlier, incomes differ from one household to the next for many reasons. Some of these differences are the result of voluntary choices made by households in the belief that they will be better off, despite having a lower measured income. Money incomes are an imperfect measure of welfare or standard of living. Redistribution based upon recorded monetary incomes may exacerbate rather than reduce inequality in standards of living. Difficulties in measuring welfare may be less relevant when a family is living in abject poverty. Measuring welfare becomes important, however, when we are contemplating policies to redistribute income among families, none of which appears to be poor. We should ensure that any redistribution undertaken delivers benefits that outweigh the efficiency and other costs imposed.

SUMMARY

Equality is an illusory goal. There is no satisfactory way of measuring inequality. Quantities that we can measure, or approximate, through statistical techniques, are poor proxies for the disparity in standard of living that is the only theoretically acceptable measure of inequality. Typically, these problems are ignored and quantities such as measured annual money incomes are simply assumed to correspond to levels of utility or welfare.

Some of the problems in measuring equality must be ignored if we are to fulfil our duty to help those living in poverty. In the case of extreme deprivation, concerns about efficiency costs, or providing assistance to those who may not really need it, give way to a pressing need to help those in distress. The case for ignoring the measurement problems is weaker, however, when we consider redistributing among people, none of whom is suffering from material hardship. We would then be incurring substantial efficiency costs from taxation and transfer policies to achieve something of dubious value.

Suppose we ignored the weak correspondence between levels of utility, or welfare, and measured annual incomes. Suppose we also ignored the efficiency costs of taxes and transfers. The goal of equality would still be unacceptable. The usual approach regards incomes in excess of the average level as equally objectionable to incomes that are below the average level by the same proportion. This only makes sense if envy of high incomes is as acceptable as compassion for those on low incomes. Envy, however, provides an unworkable basis for public policy, just as it is usually regarded as an unacceptable basis for ethical personal behaviour. (The Ten Commandments, for example, advise against it.) Much of the rhetoric used to support redistribution appeals to compassion as the motive. The measures of inequality upon which scholars rely suggest, however, that in reality much of the support for redistributive programmes arises from envy. Even though measurement problems also affect policies that can legitimately be seen as a public expression of compassion for the poor, discussion of an appropriate definition of poverty would focus attention on the relevant issue.

As we argue in Chapter 6, policies aimed at assisting the poor are likely to have many unintended consequences that compromise the possibility of achieving their ostensible goals. These programmes are likely to have other consequences that, although intended, are nevertheless not based on a compassionate concern for people in dire circumstances. Although there are good reasons for allowing or even encouraging governments to implement policies based on a compassionate concern for the poor and disadvantaged, there are also strong grounds for limiting the scope of such policies.

APPENDIX: THE GINI COEFFICIENT IN ALGEBRA

In a society of n people with incomes $y_1, y_2, \ldots, y_n$ the Gini coefficient is one minus twice the sum of the areas of triangles and rectangles with sides that have lengths $1/n$ and ratios of each income y_i to the sum of all incomes, $n\bar{y}$. This quantity can be shown to equal (Sen, 1997: 31):

$$G = \frac{1}{2\bar{y}}\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|y_i - y_j|\right] \qquad \text{(EQ 5.11)}$$

$$G = 1 + \frac{1}{n} - \frac{2}{n^2}\left[\frac{y_1}{\bar{y}} + 2\frac{y_2}{\bar{y}} + \ldots + n\frac{y_n}{\bar{y}}\right] \text{ for } y_1 \geq y_2 \geq \ldots \geq y_n \qquad \text{(EQ 5.12)}$$

and

$$G = 1 - \frac{1}{\bar{y}}\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} min(y_i, y_j)\right]. \qquad \text{(EQ 5.13)}$$

As an example, in our imaginary economy with five people earning 0, 2, 5, 7 and 11, and recalling that $\bar{y} = 5$, equation (5.11) implies

$$G = \frac{1}{10}\left[\frac{1}{25}(|0-2| + \ldots + |0-11| + |2-0| + \ldots + |11-7|)\right] = 0.432 \qquad \text{(EQ 5.14)}$$

equation (5.12) becomes

$$G = 1 + \frac{1}{5} - \frac{2}{25}\left(\frac{11}{5} + 2\frac{7}{5} + 3\frac{5}{5} + 4\frac{2}{5} + 5\frac{0}{5}\right) = \frac{54}{125} = 0.432 \qquad \text{(EQ 5.15)}$$

and equation (5.13) becomes

$$G = 1 - \frac{1}{5}\left\{\frac{1}{25}[5(2) + 3(5) + 2 + 5 + 2(7) + 2 + 5 + 7 + 11]\right\} = 0.432\,. \qquad \text{(EQ 5.16)}$$

Equation (5.11) says that the Gini coefficient can be interpreted as the average absolute difference in incomes expressed as a proportion of the mean income $\bar{y}$. Where the coefficient of variation makes the measure of inequality a function of the squared deviations from $\bar{y}$, the Gini coefficient focuses on the absolute income differences. As with the coefficient of variation, however, a proportional increase in all incomes will leave the Gini coefficient unchanged. The Gini coefficient is thus also a measure of relative incomes rather than a measure reflecting a concern that people do not fall below a minimum acceptable standard of living.

The emphasis of the Gini coefficient on income ranking is made particularly clear by equation (5.12). As Sen (1997: 32) notes, equation (5.12) implies that the Gini coefficient is a weighted sum of each person's income, expressed as a proportion of the mean, with the weights determined by the rank order of each person in a ranking by income level. Thus, the effect on the Gini coefficient of a transfer of income from one person to another (without any loss in total income) will depend on the relative income rank orderings of the two people but not on their actual incomes.

The dependence of the Gini coefficient on the rank ordering of incomes gives it a curious property. Suppose additional people are added to the society, all with incomes between, let us say, y_{i-1} and y_i. Then the *relative* weights applying to incomes $y_j - y_{i-1}$ are left unchanged while the relative weights on *all* incomes $y_j - y_i$ will change. For example, suppose $n = 10$ initially and five more people are added with incomes between y_4 and y_5 in the original population. Both before and after the new people are added, the second highest income gets double the weight of the highest income in forming the Gini coefficient. The lowest income goes, however, from having a weight of 10/9 of the second lowest income to having a weight of 15/14 relative to the second lowest income.

There would seem to be little justification for making one's valuation of the lowest income relative to the second lowest income depend on the number of people who are earning more income than either of these people. Perhaps the introduction of more people within the distribution might be seen as altering the implicit obligations of the well-off by changing the number of people who need assistance or the number of people who are able to help them. We would argue, however, that if compassion for people with insufficient access to material resources motivated one's concern for inequality, the relative concern for individuals with the lowest and second lowest incomes ought to depend more on how close each person is to genuine poverty than on the number of people who are better-off than are the two individuals in question.

Equation (5.13) suggests another interpretation of the Gini coefficient. Suppose we say that, for any pair of people i and j, the 'social value' of their incomes is just one minus the lower income of the two expressed as a fraction of mean income $\bar{y}$. Then the Gini coefficient is just the average of all these social values. For any pair of incomes y_i and y_j, therefore, only an increase in the smaller of the two incomes can raise the Gini coefficient. In this sense, the Gini coefficient is related to the maximin criterion of Rawls discussed in the appendix to Chapter 3.

6

Transfers and Taxes

After the Republican Party gained control of the United States House of Representatives in 1994 for the first time in almost half a century, President Clinton declared in the subsequent State of the Union address that "the era of big government is over". As an economic doctrine, socialism has been in retreat in the last two decades of the twentieth century. One now rarely hears the assertion that government ownership of firms is preferable to private ownership, or that centralised command and control is preferable to decentralised market-based processes, on the grounds that the socialist policies are likely to be more efficient.

The Thatcher administration in the United Kingdom, the Reagan administration in the United States, and the new policy directions of the subsequent Labour Party and Democratic Party administrations helped reinforce the idea that socialism is passé. In both New Zealand and Australia, the shifts toward more liberal economic policies were initiated by Labour Party governments.

The most potent force behind the rejection of socialism was accumulating evidence that extensive government direction of economic decisions imposes enormous costs. Examples such as North versus South Korea, or East versus West Germany, convinced many, including the leaders of China, that a greater reliance on decentralised market processes would enhance economic prosperity. The evidence was reinforced by the dismal state of the economies of Eastern Europe following the collapse of their communist regimes.

Within the developed economies of the West, evidence of the inefficiency of large government-owned enterprises gradually accumulated throughout the second half of the century. Dramatic efficiency gains following privatisation of nationalised firms illustrated the inadequacies of government ownership. (See, for example, Bishop and Kay (1988), Boardman and Vining (1989), Shirley and Nellis (1991), Kikeri, Nellis and Shirley (1992), Galal *et al* (1994) and Megginson *et al* (1994) for some of the evidence.)

The loss of confidence in Keynesian prescriptions for activist monetary and fiscal policy in the 1970s and 1980s also played a part in the demise of socialism as an economic doctrine. Far from government spending being seen as essential to the maintenance of economic prosperity, it came to be seen as part of the problem. Abandonment of traditional strictures against deficit spending except in time of war had resulted in decades of budget deficits, accumulating interest burdens and inflation.

INCOME SUPPORT

While the idea of socialism as an engine of economic efficiency may be in retreat, the idea that government ought to intervene to promote greater equality remains alive and well. The growth in government expenditure in countries belonging to the Organisation for Economic Cooperation and Development (OECD) since 1960 has been dominated by growth in transfer payments (see, for example, Tanzi and Schuknecht (1995)). Tanner and Kopel (1997) observe with regard to the United States:

> Since the War on Poverty began in 1965, federal, state, and local governments have spent more than $5.4 trillion fighting poverty in this country. How much money is $5.4 trillion? It is 70 percent more than it cost to fight World War II. For $5.4 trillion you could purchase the assets of the entire Fortune 100 corporations and all of the farmland in the United States.

The expansion in transfer payments continues even if in some cases, such as New Zealand, the growth rate appears to have levelled off relative to the growth in gross domestic product (GDP) (see, for example, Cox, 1998, Table 2.4). There is little evidence, however, that policies to promote equality have been any more successful than government intervention in the economy. Many of the social problems that motivated government intervention to address inequality appear to persist regardless of how much is spent on them. As we argued in Chapter 5 (page 148), measured poverty may have persisted partly because the poverty level has been raised by too great an extent as the economy has grown. Even allowing for problems in measuring poverty, however, genuine cases of destitution have persisted in the face of an enormous expenditure on anti-poverty programmes. The usual response to failure of government anti-poverty programmes is a call for more money to be spent. An alternative conclusion is that the programmes are fundamentally flawed.

The New Zealand government spending programmes most obviously aimed at reducing inequality include the community wage benefit (a replacement for the former unemployment and sickness benefits) and invalids and domestic purposes benefits. In addition, Cox (1998, Table 2.1) lists family support payments from the Inland Revenue Department and periodic income maintenance payments from the Accident Rehabilitation and Compensation Insurance Corporation as major benefits for persons of working age. The major purpose of a number of benefits for retired people, such as New Zealand Superannuation and transitional retirement and veterans' benefits, is to reduce the material deprivation of older people who have not saved enough to provide for their needs in retirement. The promotion of equality is a major justification for policies such as public education and public health care.

While many redistribution policies are justified on the grounds that they reduce inequality, in practice many of the benefits accrue to people who are not poor. For example, Figure 3.7 in Statistics New Zealand (1999: 43) shows that the fourth market income decile received the largest share of benefit income in New Zealand in 1996. Even the fifth market income decile received over 10% of benefit income in 1996. As we noted in Chapter 2, politicians can gain support from a majority by appearing to provide them with benefits. There is therefore an inevitable tendency for income transfers to reach families higher up the income distribution where most voters are found. Income transfers are plagued by other problems that reflect inherent defects in majority rule as a social decision-making mechanism.

Private versus government transfers

A body of statistical evidence suggests that government transfer programmes tend to displace private welfare. People are less inclined to donate time or money to private charities when they believe assistance is available from transfers financed by taxation.

For example, Roberts (1984) develops a model of private charitable giving which predicts that, in a world in which the efficiency costs of both private and public income transfers are negligible, a dollar increase in government assistance will be completely offset by a dollar decline in private transfers until private transfers are zero. Thus:

> [The model] predicts that private charity first became negligible when government first intervened in a significant way in the charity market. Significant government intervention began in the 1930s and has continued to grow over time.
>
> From the data … a stylized picture of the 1930s emerges: private donations fell dramatically as public transfers rose. But they did not fall to zero. Instead, charitable donations underwent a fundamental transformation during the period. They became less concerned with poverty and more concerned with health and social counselling …
>
> The New York Association for Improving the Condition of the Poor (AICP) was founded in 1843. For 96 years it transferred resources to the poor, the unemployed, the sick, the elderly, and the husbandless mother. Between 1843 and 1939 it was one of the two most important private charities in the city, the other being the Charity Organization Society, an umbrella for a number of smaller

> charities. Between the 1920s and 1930s the AICP's role and resources changed dramatically …
>
> With the coming of the Depression, donations, total expenditures, and expenditures on material relief increased up until 1932. This is the same pattern found in the national data. From 1932 to 1935, the experience of the AICP continues to parallel the national data – there is a decrease in donations and expenditures. But the most dramatic changes are from 1936 to 1938 when national data are missing. Donations fall to one-third of their 1932 level and two-thirds of their level in 1928. (Roberts, 1984: 142–144)

Roberts (1984: 144) notes that, by comparison:

> … in the depression of 1890–94 … public transfers in the city of New York grew only [by] 15 percent. At the same time, expenditure by the AICP grew by a factor of 4 over the 4-year period.

Furthermore:

> Even between 1929 and 1932 … private expenditures grew sevenfold nationally. Donations to the AICP continued to grow until 1932 … Only when public spending continued to grow did private spending go to virtually zero.

Roberts (1984: 144–145) quotes from the 1935–36 annual report of the AICP:

> The AICP has made major revisions of budgets downward. Many families formerly cared for by the AICP have been turned over completely to public relief departments. Nearly one-third of the present number of families under care are cooperative cases with public authorities, in which the cooperation consists in the AICP supplying social services not yet available in public departments.

Roberts (1984: 145) notes, "Additional evidence suggests that this trend was a national one and that eventually private agencies came to ignore the poor, with counselling resources going to higher-income families".

Other studies have found a decline in private expenditure on charity of from zero to 30 cents for each dollar of increase in government welfare expenditures. Roberts observes, however, that these studies have used income tax data. Sacramental donations to religious institutions (that do not go to the poor) make up about half of these contributions, while educational and medical institutions receive about 15% each. The social services category receives about 10% of tax deductible donations. The social services category includes agencies such as the Salvation Army, religious organisations such as the United Jewish Appeal, Catholic charities and the Federation of Protestant Welfare Agencies, the Scouts, and the Young Men's Christian Association (YMCA) and Young Women's Christian Association (YWCA). About one third of the social services category is received by the United Way. Roberts (1984: 141) observes:

United Way spending varies by community but typically includes spending on health and other items that only partially help the poor, such as day-care centres.

There is another problem with studies based on income tax data. An increase in government spending that is accompanied by a rise in marginal tax rates will reduce the private cost of making eligible contributions to exempt organisations. We agree with Roberts that the evidence points to a relatively high degree of substitutability between government and private income transfers to the poor.

Government welfare payments affect more than the willingness to provide charitable contributions to help the poor. Transfers affect the incentive to work, attitudes to marriage and the decision to have children out of wedlock or the decision of fathers to support their children.

For example, Tanner and Kopel (1997) examined major studies based on US data that were done by people ranging from mainstream labour economists such as Mark Rosenzweig, Robert Moffitt, Robert Haveman and Shelley Lundberg, through to employees of the US Department of Health and Human Services. Most of these studies reveal that a 10% increase in welfare benefits results in an average 10% increase in births to unmarried, low-income women.

The studies Tanner and Kopel (1997) examined also found that, in the United States, welfare payments had a statistically significant negative impact on labour supply and hours of work among recipients. Welfare payments also had a statistically significant negative impact on labour force participation of young men in communities with high levels of welfare participation, even though the young men were usually not themselves direct recipients of welfare benefits.

Labour market studies have found that prior work experience has a major positive effect on current wage rates. People who respond to welfare benefits by avoiding work when young therefore are likely to suffer a lifetime of reduced earnings. A lack of work experience also negatively affects people's self-esteem and the likely future prospects of their children.

Income transfers also influence whether people take out market insurance to offset the consequences of adverse events, as well as saving as a form of self-insurance against adverse reductions in income or increases in expenditure. Furthermore, family members and friends who are in need are expected to seek assistance from government agencies. Transfers to the young and old, the sick, the unemployed or the injured, sole parent allowances, marriage tax penalties, and publicly funded childcare tend to weaken family ties and displace the traditional role of families in assisting their own relatives who are in need. Government transfer programmes may, however, be a poor substitute for the private measures that they partially displace.

Families and friends, in particular, are likely to be substantially better than government at providing temporary support to people in need. People in temporary difficulties often need emotional and moral support as much as material support. Such support can only be offered by families and friends. Families also perform other services, such as transmitting cultural values, providing general education and improving work habits, while rendering material assistance. When government transfer programmes relieve people of their obligation to assist others in need, the government may be incapable of replacing the non-material support that is withdrawn.

Families and friends are also likely to have much better information about a recipient's needs, tastes and circumstances. They can tailor support to match that information. The government welfare system is of necessity large, standardised and relatively impersonal.

Families and friends may also be better able to control moral hazard on the part of recipients. As we discussed in Chapter 4, moral hazard will be a major problem for anyone providing material support to someone else. Income support tends to reduce the incentive to avoid poverty or to reestablish independence after misfortune. Families and friends will in general have good information about the steps that recipients have taken to avoid their plight, or steps they have taken to improve their situation. Families and friends also have recourse to effective non-pecuniary penalties, such as ostracism, for individuals who exploit the generosity of others. Access to effective non-pecuniary penalties is especially important in this context, since it may be counterproductive to impose pecuniary penalties while trying to help someone who is suffering from material deprivation.

While reliance on families and friends has many advantages relative to government income transfers, it also suffers from one major defect. Some people who find themselves in desperate circumstances may have no friends or family to whom they can turn for assistance. Voluntary assistance and donations to religious and other charities can greatly assist these people. There are good reasons to believe that private charitable organisations are likely to be more successful than governments in providing such assistance.

We noted in Chapter 2 that governments typically have much less information than do private individuals taking part in voluntary transactions. Some of the information governments believe they have is also inaccurate as a result of distorted incentives:

> As the social distance between the donor and the recipient increases, it becomes more difficult to pick out deserving recipients from operators milking the system for all it's worth. Nor is it easy to determine either the proper form or the amount of the transfer … Even advocates of voluntary assistance programmes worry about their incentive effects, and constantly look for levels and forms of support sufficient to benefit those in need without undermining their willingness to do productive labour …
>
> The difficulties that beset voluntary redistribution are largely self-correcting, for if a programme fails consistently, its champions eventually abandon it. The precarious nature of charitable giving is one of the features that make it most effective …
>
> The systems of public welfare are in many instances unable to perform nearly as well as the private systems that they displace. Taxpayer cries for supervision of public officials and their charges rule out the informal systems of control available to many charitable endeavours. Instead, public welfare programmes must rely on impersonal bureaucrats, always subject to norms and proofs, which are utterly impervious to undocumented forms of common knowledge and common sense, no matter how reliable. (Epstein, 1995: 142–143)

A related problem is the difficulty governments have in monitoring the performance of government employees and agencies. There is no readily available source of information on agency performance, such as a stock price. Monitoring problems are generally exacerbated when agencies are providing a wide range of services. A consequence is that governments typically have to provide a limited range of options in support programmes. There is a tendency to use a one-size-fits-all approach.

By contrast, private charities typically have fewer clients with whom to deal and are often closer to their clients. Private charities can, therefore, better tailor their support to an individual's circumstances. Private organisations also often cater to specific subgroups of the population that tend to be homogeneous and have similar needs. For example, in the United Kingdom, Australia, New Zealand, Canada and the United States, fraternal organisations and friendly societies provided life, sickness, disability and unemployment insurance to their members throughout the second half of the nineteenth and first third of the twentieth centuries. These organisations tended to serve people in homogeneous demographic, ethnic, religious or occupational groups. Since the people came from similar backgrounds, and were often known to each other socially, the societies found it much easier to control moral hazard problems than do government agencies today.

By providing charitable support along with other services (such as religious instruction), private organisations often increased the homogeneity of their clients. Green (1996: 68) notes that the activities accompanying the support produce another significant difference between private and public assistance:

> [T]he very act of paying or not paying a benefit sends a signal of approval or disapproval. When the state began to send cash to unmarried mothers to raise their own children instead of sending a social worker to encourage them to have their babies adopted by other couples able to give them a better chance, the moral message changed. Public policies convey moral messages in addition to producing incentive effects. Good conduct, according to Wilson and Herrnstein, is based on "internalised inhibitions, reinforced by social sanctions" and public policies are among the signals of approval or disapproval being transmitted.

Competition among agencies for donor funds and staff will tend to result in many different types of ancillary services being offered to clients. Experimentation with different ways of doing things is encouraged, while agencies can learn from the experiences of each other. In a system in which reduced government activity allows many private alternatives to flourish, people in need are more likely to find an organisation with which they are comfortable. Surveying the types of organisations currently available gives a distorted picture of what is likely to happen in a deregulated, competitive environment.

The conditions private agencies impose in return for providing assistance may cause some individuals to forgo the offer of help. Coverage may be less complete than a government programme providing support without any obligations on the part of the recipient. As much as possible, however, the government ought to be a provider of last resort. People in need should be encouraged to rely on personal savings, private insurance arrangements, family, friends and private charity. Public assistance ought to be at a level that allows private assistance to do most of the work.

Subsidy versus government provision

Even if we conclude that, despite the potential adverse consequences, there is an argument for government transfers to supplement private activity, private entities could supply the services under contract to government agencies. The common bias toward monopoly public provision may reflect the tendency for democratic political processes to be captured by vested interests, rather than any inherent advantage in government supply.

Government agencies are more difficult to monitor than private firms that have to meet public accounting and reporting standards. In particular, share prices provide a continuous market assessment of the quality of the decisions made by managers of private firms. Investors also have strong incentives to monitor management decisions. If current managers are not using the firm's assets efficiently, a majority shareholder can reap a capital gain by installing new managers who will make better decisions. Competition among private firms for customers also encourages managers to control costs. Monopoly government agencies face few of these pressures for increased efficiency. Politicians also have much weaker incentives than investors in private firms to monitor the decisions made by managers of government agencies. There is at best a weak link between a politician's reelection chance and excessive costs of operating a government agency. The weak control over government agencies means, however, that employees often prefer to work in such establishments. Private organisations tend to have stronger performance requirements and stricter controls.

Expenditures on public education, public health care and other services ostensibly provided to the poor are at least partially captured by the suppliers. Indeed, the tendency to provide services such as counselling, education, medical care, childcare and so on as opposed to cash transfers can be seen in part as reflecting the vested interests of the suppliers of such services.

The failure of such programmes can be seen in public education. Many have argued in favour of assisting poor families in educating their children because it is equitable to do so. Since many economic studies have shown that education generally increases future income, education may be an important path to intergenerational mobility. Even though education can greatly increase future income, however, a family without physical collateral may find it difficult to borrow money to pay for the tuition and living expenses of a child who is in school.

There is also evidence that education provides economic benefits beyond its effects on labour market productivity. For example, we noted in Chapter 5 (page 115) that education may help people gain more value from the same level of expenditure. These benefits, however, may not show up as a greater ability to service future loans, and thus may not allow a person to borrow to finance the educational investment.

Since a private bank has no easy way of forcing a borrower to pay back such a loan, it is unlikely to lend money for educational investments. The government can, by contrast, use taxation to enforce loan repayment, although, in practice, the default rate on government-guaranteed educational loans in the United States is quite high.

Even if people have difficulty borrowing to finance educational investments, such loans should not be subsidised unless the educational investments provide spill-over benefits to others. In most cases, therefore, the loan obligations ought to be enforced.

Some have suggested that there may be efficiency arguments for government educational grants to the poor, or for subsidising educational loans. As we noted in Chapter 4 (page 103) there may be spill-over benefits associated with a basic level of education. Some of the poorest families may voluntarily purchase a level of education for their children that does not take these benefits into account. Subsidies might then increase efficiency by raising education expenditure to the point at which these spill-over benefits are zero. Education beyond this level would primarily benefit the person being educated, and the spill-over benefits of additional education will be zero.

Furthermore, there are many private scholarships available to assist poor families to educate their children. There would probably be more of these if the government were not so extensively involved in education. For example, private universities in the United States offer substantial financial aid to children from poor families.

As West (1965: 51) has noted, it is also possible to overstate the benefits of education as a means of increasing equality:

> A monopoly labour union which forces a closed shop upon its employer and then rations the number of union cards available creates a much more obvious barrier to workers or to school-leavers outside the union than does any deficiency in education. As another example, a 16-year-old boy has no equality of opportunity at all with a boy of 15 years to enter certain trades if the rules of apprenticeship decree that he is too old. Moreover, we cannot always be sure that long programmes of work-based education (including, for example, day-release at a technical college) which await 15-year-old successful candidates for apprenticeships have not been designed with some view to discouraging the numbers in the trade in order to raise the earnings of those already in it. Where this is so, further technical education is not a key to equality of opportunity but one of the means of suppressing it.

West (1965: 58) also makes the point that educational attainment depends to a large degree on native intelligence, and abilities such as intelligence are not equally distributed in the population:

> Suppose that we can counteract all the subtle environmental handicaps by some additional and intensive education for the naturally dull children sufficient to get them up to the starting line of the race. Suppose also that beyond this we can even prevent some parents spending extra amounts on their children merely by virtue

> of their being richer than others. It is quite obvious that it is impossible now for all competitors to finish together. In other words where there is inequality of potential (ability) there is bound to be inequality of result. If we insist that there shall be equality of result it follows that we penalise ability.

Most importantly, however, one can support government financial assistance for education without requiring that educational services be supplied by a monopoly government agency. Tax revenue can be used to subsidise competitive private schooling for the poor through means-tested vouchers. Such vouchers would increase equality of access to education, because poor children could attend schools that are now available only to the rich. (For an extensive discussion of the benefits of promoting competition in the provision of education services see Green, 1996: Chapter 8.)

As we noted in Chapter 3, there are many problems associated with representative democracy. Government provision of education illustrates some of them.

Government monopolisation of the provision of educational services has had similar consequences around the world. The size of bureaucracies, and the ratio of administrators to teachers, has risen inexorably. Educational reforms that make life easier or more comfortable for the administrators or teachers tend to win acceptance. Unfortunately, learning is hard work, and it places demands on teachers as well as pupils. There is pressure to eliminate objective measures of student performance, since such measures also compel schools to produce better educational outcomes. There is little or no competitive pressure for the schools to provide a service that is valued by parents. Indeed, government administrators who provide funds, not parents, are seen as the customers. Complaints about the declining quality of education are invariably met by demands for additional funding rather than elimination of the past round of reforms, more competition between suppliers, or more effective monitoring of teachers.

Public education also illustrates the tendency of majority rule to expand transfer programmes to encompass the middle class. Most public school students, as well as the majority of voters, are from middle-class families. It therefore would appear to be easy to explain why politicians can garner majority support by favouring public education. Since the public education monopoly is much less efficient than the competitive system it displaces, however, the extra taxes that middle-class families pay over their lifetimes exceed any savings they accrue from not having to pay for competitive private education. Majority support for public education is thus more paradoxical than it seems.

Perhaps voters are not aware of the additional tax burden resulting from the public education monopoly. Even if voters are aware of this, however, it may be rational for many of them to favour public education. The education expenditure is not tied to any particular source of tax revenue, so a reduction in expenditure on public education may not result in lower taxes. At least some of the revenue is likely instead to fund increases in other public expenditure programmes. As we remarked in Chapter 3, voters have an incentive to lobby for their particular interests, in this case their interest in using public education services, and to pay less attention to their general interests as a taxpayer.

The short time horizon of politics

Frequent elections force politicians to focus on the short term and on obvious effects of policies. This is not to say that we would be better off with fewer, or even no elections. Periodic elections are required to provide an incentive for the government to respond to the wishes of voters. A defect of the decision-making mechanism, however, is that it focuses attention on the immediate and obvious effect of policy at the expense of the distant and indirect effects.

The tendency to focus on short-term effects has a number of undesirable implications for policies to address poverty. Poverty is often viewed as a static phenomenon. On this view there is a certain group of people who are poor and who always will be poor. Economic studies have shown, however, that there is substantial turnover in the people who are found below the poverty line from one year to the next. Some of these studies are discussed in more detail in Chapter 7.

The dynamic nature of poverty makes some policy responses particularly inappropriate. For example, a public housing policy may allow people to qualify for assistance when they are temporarily poor and then keep the subsidised housing when their income subsequently improves. One consequence is that public housing tends to be occupied by many families who do not have a low income.

Fortunately, most housing assistance in New Zealand is now provided through the accommodation supplement which provides a subsidy for renters (and borrowers) on low incomes whether they rent from private or public landlords. People who rent from Housing New Zealand Limited (a Crown entity) are supposed to pay a market rent (before any supplement). Apparently, some former Housing Corporation rental units were provided on a subsidised basis to people who were not poor. A reason for privatising

Housing New Zealand Limited is to reduce the chance of reinstituting such a policy. There are no compelling reasons for the government to provide rental accommodation for low-income families. Such housing is a private service that can be supplied through the market.

While poverty is, for many, a temporary phenomenon, there are still people who remain poor, and dependent on public income support, for substantial periods of time. Cox (1998: 24) cites a briefing paper by the Treasury to the incoming government in 1996:

> [T]he briefing notes that, as at June 1996, 54 percent of working age beneficiaries had been in receipt of benefit for more than a year and that 109,000 children in March 1996 lived in families that had been continuously in receipt of benefit for two or more years (p 82). The average duration of receipt of domestic purposes benefit had increased from three years in 1982 to three and three quarter years in 1996. Moreover, although 'around 40 percent of [DPB] benefit recipients move off benefits within 12 months of entry, many of these do not stay off. Nearly 60 percent of DPB recipients who exited the benefit in the second half of 1992 returned at least once within the three years to June 1995. As at June 1996, nearly a quarter of the total DPB population had continuously received the DPB for at least six years' (p 87).

In order to succeed, politicians must appear to assist people in temporary need. There is little incentive to ensure that the people assisted are helped more than they are hurt, and even less to ensure that the long-term consequences are favourable. The notion that government programmes will affect incentives is mostly absent from the public debate.

Income transfer programmes also have a number of adverse indirect effects that are not widely discussed and have little influence on the decisions of the average voter on election day. More generous welfare programmes require higher levels of taxation to finance them. If the number of people receiving benefits is to be kept under control (in order to limit the cost), higher benefit levels lead to higher abatement rates for means tests or extend the income range over which benefits are phased out. Higher tax and abatement rates further discourage participation in market activity and exacerbate the direct problems of transfers discussed above.

In many countries, the increase in government expenditure on income transfers has been accompanied by decreases in government expenditure on public infrastructure, including law enforcement, as taxpayers have resisted escalating tax rates. The adverse consequences of such expenditure reductions may fall disproportionately on the low-income segment of society, as they are less able to purchase the privately provided substitutes.

Chronic government budget deficits arising from increased income transfers have also raised the cost of private investment. Lower investment in turn reduces economic growth below what it would otherwise have been and limits the increase in labour productivity.

The perverse incentive effects of income transfers, and the indirect effects of higher taxes, reduced expenditure on public infrastructure or larger budget benefits, may all reduce opportunities for future generations. The attempt to reduce inequality at a given moment in time through compulsory taxes and transfers may actually reduce intergenerational mobility. Again, however, the incentives are weak for politicians, or civil servants, to consider these long-term effects.

The influence of other shared beliefs

As we discussed in Chapter 3, not all beliefs that are widely held are ethically sound, and majority support for a policy is no guarantee that the policy will be a good one. For example, some voters oppose any income transfers because they believe the poor are in their situation through their own fault, or that most recipients come from an identifiable racial or cultural subgroup of the population.

On the other hand, many supporters of redistribution object to transferring income, as opposed to in-kind services. In particular, critics of market economies harbour a deep-seated hostility to the way the market responds to the wishes of the consuming public. These critics often deride popular tastes and believe they know what is best for others. They are opposed to transferring income and to allowing recipients to spend it as they please. The implicit constraints on individual freedom would be considered obnoxious if imposed on higher-income people. In some cases, supporters of in-kind as opposed to cash transfers supply the subsidised services, so their opposition to cash transfers may not be genuine. The issue of cash versus in-kind transfers was discussed in more detail in Chapter 4.

PROGRESSIVE TAXATION

Views about equity also play a prominent role in public discussion of taxation policy. A compassionate concern for people who are poor can be advanced much more effectively, however, through income transfers than through tax policy. At best, taxes could leave low-income earners with the same income. The issues of incentives and efficiency that have been emphasised by economists are a more appropriate basis for evaluating different systems of taxation.

As we argued in Chapter 4, the efficiency losses from taxation are likely to increase more than proportionately with the tax rate. Since higher tax rates reduce incentives to produce income, and increase incentives to evade and avoid taxes, increases in tax rates may not even raise much revenue. Nevertheless, many commentators and politicians continue to call for increases in marginal tax rates applying to higher reported income levels.

It is often assumed that an income tax system which is progressive in the marginal tax rate (a 'graduated tax') is more equitable than one in which the marginal tax rate is constant. Progressive marginal tax rates are more equitable than constant marginal tax rates, however, only if one is motivated by envy with regard to high incomes. If one is not motivated by envy, a tax with a constant marginal rate, and an income exemption, could be more equitable, as well as more efficient. This is not to say that an income exemption is desirable on efficiency or equity grounds. As we shall see, it may be better to eliminate the income exemption and instead provide transfers to those living in poverty through the welfare system.

Average versus marginal tax rates

An income tax system is considered progressive if individuals who have higher pre-tax incomes pay a higher proportion of their income in taxes. It is often thought that progressivity requires that the marginal tax rate, which is the amount of tax paid on the last dollar earned, rises with income. This is not correct. The tax system will also be progressive if there is a constant marginal tax rate, combined with an exemption level of income below which no tax is payable. Such a system is said to be linearly progressive.

Figure 6.1 graphs total tax paid as a function of income earned under a linearly progressive tax system. The proportion of total income paid as taxes is the average tax rate. It is represented in Figure 6.1 by the slope of a line from the origin to the total tax payment schedule. As a result of the income exemption level, it is clear that this slope rises with income and thus that a linear system can indeed be progressive. For example, the figure includes the slopes for a taxpayer earning $25,000 and another earning $55,000. With a flat marginal rate of 33%, the tax payable on $25,000 after subtracting the $10,000 exemption is $5,000, resulting in an average tax rate of 5/25 = 20%. Similarly, the tax payable on $55,000 is $15,000, resulting in an average tax rate of about 27.3%. As incomes continue to increase, the average tax rate approaches the marginal rate of 33%, since the $10,000 exemption comes to mean less in relative terms.

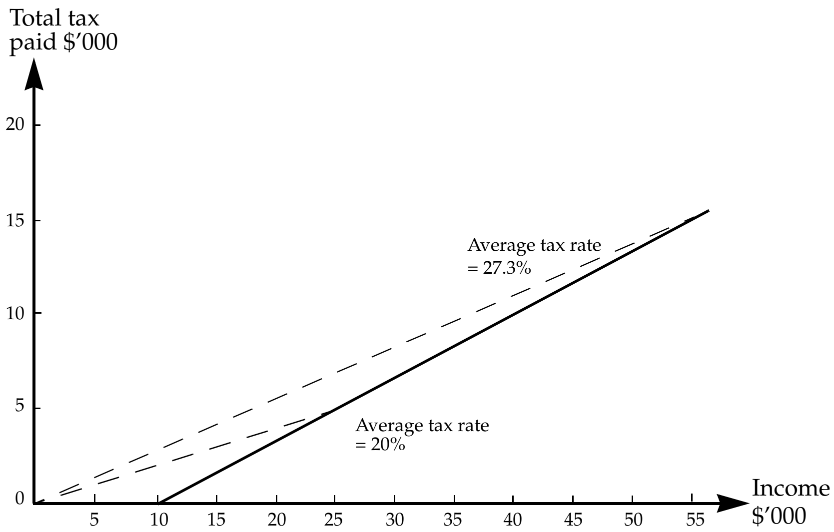

FIGURE 6.1: A linearly progressive income tax

Efficiency costs and tax schedules

Since people know that their personal shares of total tax receipts are small, each person would prefer to pay as little tax as possible. In consequence, taxation leads to several types of distortions, or sources of efficiency losses. Some of these distortions depend on the marginal tax rate, while some depend to a greater extent on the average tax rate.

Tax can be avoided by obtaining utility in a non-taxable form. Employees may take some compensation in the form of untaxed fringe benefits, better working conditions or a more pleasant work environment. The New Zealand government taxes major fringe benefits, such as the provision of a motor vehicle for private use and low-interest loans, but other benefits remain tax-exempt. Even if the benefits of these amenities to employees are not as great as the cost to their employer, the lower taxation of such benefits may make them preferable to wage income. The result, however, is an inefficient use of resources, a loss of tax revenue and higher rates of tax than otherwise on other income or goods and services.

The incentive to avoid taxes through fringe benefits depends largely on the marginal tax rate. Suppose someone is deciding whether to take an extra dollar of income as an additional fringe benefit, rather than as taxable money income, or whether to get extra furniture in the office or better food in the office cafeteria. The tax applying to an additional dollar of money income will affect how the person decides to spend the extra dollar.

Taxation also encourages people to work less, produce less money income and consume fewer market goods and services. People increase their utility instead by working around the home, attending sporting events, and engaging in other activities that involve lower expenditures of income but more direct time input. Again, resource use is likely to be inefficient, since the resource cost of providing more market goods and services will fall short of the benefit of doing so. The reduction in market work and the consumption of market goods and services will also eliminate tax revenue the government could have raised in the absence of such disincentive effects.

Both marginal and average tax rates are likely to affect incentives to work. On the one hand, people thinking of earning more income through overtime, or making other decisions to raise their taxable income by a relatively small amount, are likely to be sensitive to the marginal tax rate. The substitution of additional market for non-market activity will respond to the after-tax income earned for each additional dollar of pre-tax income. On the other hand, an increase in the average tax rate will have a negative income (or impoverishment) effect. An individual would have to supply more hours of work in order to attain the same level of consumption of market goods and services. In order to maintain consumption as the average tax rate rises, the person will, therefore, be inclined to work more hours.

It might be thought that the efficiency costs of high taxes, in addition to their disincentive effects, will be muted by the offsetting negative income effect. One has to be careful in applying this result, however.

The revenue raised through taxation does not simply disappear from the economy but instead is returned to people in one form or another. Thus, a higher average rate of tax for everyone would be associated with a higher average level of government expenditure, and these goods and services will have an opposite income effect on hours of work. If people value the goods and services supplied by the government as highly as they value market goods and services, government expenditure will tend to reduce hours of work.

More importantly, the efficiency effects of taxes result from the distortion of choices when marginal benefits or costs are artificially made to look different from the true marginal benefits or costs. The size of these losses depends on the slopes of the marginal benefit and marginal cost curves, as a function of the level of activity undertaken, and income changes have no effect on these slopes. As we noted when we introduced the concept in Chapter 2, the efficiency criterion essentially ignores all income effects by treating a dollar of income the same regardless of who earns it.

Where a capital gains tax is applied, people can often delay tax payments by delaying the realisation of taxable gains on investments. This can be seen by contrasting two investment strategies for a given project. One strategy is to realise all taxable gains each period as income, pay tax on that income and then reinvest the after-tax proceeds into the project. The other strategy leaves the pre-tax gains unrealised. The latter strategy will accumulate more value in the project and produce a larger taxable gain when the project is finally liquidated. The tax that is paid under the first strategy, but appears to be avoided by the second, is apparently paid upon liquidation. There is, however, a timing advantage of the second strategy. By delaying the payment of tax, a person following the second strategy ends up incurring less tax in present value terms. The taxpayer, in effect, obtains an interest-free loan from the government equal to the accumulated value of the interim tax liabilities. New Zealand has eliminated these problems for some financial arrangements. Certain financial assets, such as zero coupon bonds, are taxed on an accruals basis. The distortions remain, however, for equities and also for real assets that can only be valued easily when they are sold.

Income taxes distort timing decisions in other ways. In particular, people have an incentive under an income tax to expense immediately expenditure that provides future benefits, and hence ought to be capitalised and depreciated over the life of the item that was purchased. For example, the income tax produces an incentive to spend funds maintaining old equipment, which may be deductible against current income, rather than scrapping the equipment and buying a new replacement, which could only be depreciated for tax purposes. In New Zealand, some investments, particularly in forestry, are allowed to be expensed while most investments must be depreciated. Disparities in tax treatments of different investments distort incentives to allocate funds to alternative projects.

Finally, tax avoidance and evasion can be viewed as two more distortions induced by taxes. People can use mechanisms, such as manipulating their accounting records and establishing overseas bank accounts, in order to hide income from the tax authorities and pay less tax.

The resources used in establishing tax avoidance mechanisms, including the efforts of many highly trained accountants and lawyers, are wasted and constitute another part of the efficiency costs of taxation. Furthermore, the money which would otherwise accrue as tax revenue is shared instead among the taxpayers, their lawyers and their accountants. Higher tax rates are needed to raise the same revenue, increasing the losses from other distortions.

The incentive to engage in accounting schemes and related measures of tax avoidance may depend most on the average tax rate. Many accounting and legal arrangements may have a fixed cost of establishment that is largely independent of the amount of income sheltered from tax. Only taxpayers with a large tax bill would find it advantageous to exploit such schemes. Furthermore, once the mechanism has been established, the amount of tax avoided or evaded may not depend much on the marginal tax rate.

The inefficiencies associated with tax distortions depend on the amount of the distortion and the difference between the costs and the benefits of the distorted activity resulting from the tax wedge (see also the discussion of the efficiency costs of taxation in Chapter 4). Since the amount of the distortion is likely to be roughly proportional to the tax wedge, the total efficiency costs will depend on the square of the tax rate. A doubling of the tax rate will double the difference between the marginal costs and marginal benefits of the distorted activity. It will also roughly double the reduction in the amount of the taxed activity that is undertaken. The efficiency losses, or the lost net benefits that could otherwise have been supplied to consumers, will therefore increase by a factor of four. The higher tax rates under a progressive tax system thus exact a disproportionately high efficiency cost.

Tax scales with multiple rates also impose high administrative and compliance costs. Taxpayers have an incentive to shift income or deductions among individuals and years to alter tax liability. Much attention is devoted to manipulating the timing and nature of expenditures or receipts.

The extent to which manipulations of expenditure or income are encouraged depends on the number of different marginal tax rates in the progressive tax system. Even a tax with a constant marginal rate above a non-trivial exemption level will invite some of the same manipulations if income can be allocated to a taxpayer with income below the exemption level in any one year.

Under a flat tax, with a single marginal and average rate of tax, income earned through entities such as companies could be taxed at the company level at the single rate of tax. Dividends could then be distributed tax-free and the imputation system dropped. Financial institutions could deduct tax on interest earned by depositors and such interest could be exempted from further taxation. Substantial numbers of taxpayers who only earn income that is taxed before it is paid, such as salaries, wages, dividends and interest, could be exempted from filing a tax return. Tax rules aimed at limiting the scope for income splitting among high- and low-rate taxpayers could be abolished. Provisions relating to income equalisation schemes could also be eliminated. These provisions currently apply to farmers earning income from forestry and farmers or agricultural businesses using income equalisation schemes to smooth income in response to adverse events. Fewer people would be needed to administer the tax system, and the expertise of tax accountants and lawyers could be more gainfully employed on other tasks. As Blum and Kalven (1978: 14–19) note:

> The price the tax system pays for progression is thus high. It produces a tax law of almost impenetrable complexity. It invites a distorting attention to the tax aspects of any economic transaction. It affords an excessive stimulus to tax avoidance with perhaps incalculable consequences for taxpayer morale and the general respect for the law.

The higher costs might be justifiable, but only if there were sufficient offsetting benefits.

Proposed justifications for a graduated tax

Many scholars have attempted to justify progression in marginal income tax rates. Such a graduated income tax has been supported on the grounds that it more nearly equates taxes paid to the benefits received from government expenditure, or more equitably apportions the sacrifice associated with paying taxes.

The most common arguments given in favour of a graduated income tax implicitly make several dubious assumptions. One assumption is that a tax on income is paid by the person upon whom it is levied. While it is true that the person legally responsible for a tax makes the payment to the government, the tax is likely to alter pre-tax incomes and prices. In consequence, consumers, or other suppliers of services to firms, may bear part, or all, of the burden of the tax.

Suppose we assume that the final economic incidence of taxes corresponds roughly to the legal incidence. Arguments concerning the justice of different income tax schedules still implicitly assume, contrary to our analysis in Chapter 5, that annual monetary incomes measure individual welfare.

Even if we grant these dubious assumptions, the case for progressive taxation has further difficulties. The most common argument made in favour of progressive taxation is that it more nearly matches tax payments to the benefits received from government expenditure than would a proportional or per capita fixed tax. This is unlikely to be true. Those who earn zero taxable income, for example, typically benefit a great deal from government-provided goods and services, including income transfers.

Some have argued that maintaining law and order, or providing defence against external threats, are the major services provided by government and these services are of much greater benefit to the wealthy (those with substantial private property in particular) than to the poor. The composition of government expenditures suggests otherwise. For example, the New Zealand Treasury Pre-election Economic and Fiscal Update (1999) reveals that, for the year ending 30 June 1999, law and order was slightly over 4% of total government expenditure and defence slightly under 3%. This compares with 36% allocated to social security and welfare, 18.3% to health, 16.5% to education and 4.8% to core government services. Second, wealth need not correspond closely to a high annual money income, and lack of property need not correspond to a low annual money income. Third, even when it comes to expenditure on law and order, protecting individual liberty and personal integrity are at least as important as protecting property. Public enforcement of laws protecting individuals against personal violence, unreasonable search and seizure, harassment, unconscionable contracts and so forth, may actually be of greater value to the poor. Finally, the wealthy often help to deter crime in many countries, including New Zealand, through security systems and private guards. Public policing may actually provide greater benefits to the law-abiding poor than to the law-abiding wealthy.

Economists have attempted to measure the likely distribution of the benefits of public expenditure. While none of these measurements is perfect, they provide some useful information. The usual result is that public expenditure redistributes from either end of the income distribution to the middle. Middle-class citizens consume most of the public expenditures on education (including at the tertiary level). Better educated middle-class citizens also benefit disproportionately from public expenditure on health care. They are often better informed than the poor about the available services, are better able to supplement the public services with private care, and are likely to live longer. Even income transfers tend to go disproportionately to the middle class, once one includes assistance for the aged. The middle class also tends to use public roads and other subsidised transport services more than the poor. Charging taxes to match the benefits of public expenditure would not lead to a graduated income tax scale. Moreover, it is not clear that taxes ought to match purported benefits of government spending, as the ethical premise implicit in this claim has not been justified.

There may be an efficiency argument for applying user charges where benefits can be attributed to particular users. Such charges could signal to users the marginal costs of meeting their demands for services. If users could moderate their demand, they then would have an incentive to equate the marginal benefits of the government service to the marginal cost. In turn, the government would obtain information about the value users place on the service that could be compared with its cost of provision.

For items of government expenditure such as law and order, and defence, however, user charges are not feasible or would not be effective at rationing demand. Taxes impose efficiency losses because they discourage some activity (for example labour supply or consumption of market goods or services) which has a marginal cost that is below its marginal benefits. By contrast, a price or user charge that appropriately signals marginal costs encourages each consumer to ration demand to a level at which the marginal benefits equal the marginal costs. Taxes distort price signals and make people misapprehend the marginal costs or benefits of taxed actions. Taxes are most appropriately used to finance public expenditure when user charges are not feasible or serve no purpose. Once all opportunities for applying user charges have been exploited, there is no efficiency gain from relating remaining taxes to estimated benefits.

Another popular argument for the graduated income tax postulates that all tax revenue is solely a sacrifice for taxpayers. This argument ignores the benefits of government spending altogether and instead focuses on how the burden ought to be shared equitably. The theory is discussed in more detail in an appendix to this chapter. The sacrifice associated with taxation is quantified by assuming a particular utility of income function. Depending upon whether the goal is to equalise the sacrifices of taxpayers, impose the same proportional sacrifice on taxpayers, or minimise the total sacrifice, one can derive different degrees of progressivity in the tax structure.

In addition to ignoring the benefits of government spending, the sacrifice theory ignores the disincentive effects of higher taxes. When we introduce disincentive effects, however, we need to make further assumptions in order to derive an implied degree of progression in tax rates. As we shall see, it is not obvious that progression in the marginal rates is the result.

A more sophisticated version of the sacrifice theory proposes that the government ought to choose the level of taxes (and the level of public expenditure) to maximise a so-called 'social welfare function' $W(U_1, U_2, U_3, \ldots, U_n)$ that depends positively on each person's utility level (so pure Pareto improving changes are valued positively no matter who gains). While choosing tax rates to maximise *W*, the government takes account of the disincentive effects of taxes. The government is assumed to know the factors that in principle determine individual utility, but not to have sufficient information to know the specific utility level of any one individual (see Mirrlees, 1971).

To allow for disincentive effects, we need to give each person a choice that will be distorted by the imposition of taxes. The obvious modification is to make the amount of market labour depend on the after-tax income. In order to allow individuals to earn different incomes while retaining the assumption that all utility functions are alike, we assume individuals differ in their earning abilities, which are reflected in market wage rates. While utility functions are alike, people will not enjoy the same utility levels because each person has a different ability to earn market income.

If the government knew each person's earning ability, it could devise taxes specific to each individual. This is an unreasonable assumption. If the government did know ability levels it could determine the specific utility of each person. The government could then devise a unique tax for each person that depended on the ability, but not necessarily the market income, of that person. Taxes would impose no efficiency costs. Furthermore, in practice, taxes can be based on actual market income but not on each person's innate ability. Thus, earning ability is assumed to be known to each individual but not known to the government.

Market wage rates depend on effort as well as earning ability and the government also cannot observe the amount of effort people are putting forth. When the government attempts to tax high-ability people more heavily, therefore, those people put forth less effort in addition to supplying less market labour. In effect, they have an incentive to pretend to be lower-ability people so they do not have to pay as much tax. While the reduction in market labour may in principle be observable (by looking at wage rates and income levels), the reduction in effort is not.

The key conclusion from these models that allow for disincentive effects of taxes is that the marginal tax rate, or the additional tax paid on an additional dollar of income, ought to be zero for the lowest-ability, but also for the highest-ability, people. The tax can include a per capita component that is independent of market income. Since this is also the same for everyone, it will not, under the present assumptions, produce any inefficiencies. Thus, the per capita component should be set to extract the amount of tax (or pay a subsidy if the tax rate on the lowest-ability people is negative) that would otherwise be raised from (or paid to) the lowest-ability person. The lowest-ability person then faces a zero rate of tax on income and chooses to supply efficient amounts of both labour time and effort.

For the highest-ability person, consider the following argument. Suppose this person is facing a positive marginal tax rate. The government could then commit to not taxing any additional income this person earns (that is, add another income bracket with a zero marginal rate). The person would have an incentive to earn more income than before, while the government would raise the same revenue it was raising from this person before introducing the new zero marginal rate. No one with a lower level of ability will be affected (this may appear obvious, but it is difficult to prove) and the result will be a Pareto improvement.

The result that a zero marginal tax rate should apply to those of highest ability prompted many analyses aimed at overturning the conclusion by modifying various assumptions. There is, however, a fundamental problem with all of the benefit-and-sacrifice arguments. Such arguments attempt to derive normative conclusions from an entirely positive analysis. A more open approach requires making the ethical principles explicit.

Ethical principles underlying graduated taxes

Suppose we have a tax system that is progressive in the marginal rates. To be concrete, suppose the tax scale is as given in Table 6.1. Consider changing this system to the linear progressive tax given in Table 6.2. The total taxes payable under the two systems are graphed as a function of income in Figure 6.2.

TABLE 6.1: A tax scale progressive in marginal rates

Taxable income	*Tax payable*
less than $10,000	$0
$10–30,000	15¢ per dollar in excess of $10,000
$30–46,000	$3,000 plus 25¢ per dollar in excess of $30,000
above $46,000	$7,000 plus 40¢ per dollar in excess of $46,000

TABLE 6.2: A linear progressive tax scale

Taxable income	*Tax payable*
less than $15,000	$0
above $15,000	25¢ per dollar in excess of $15,000

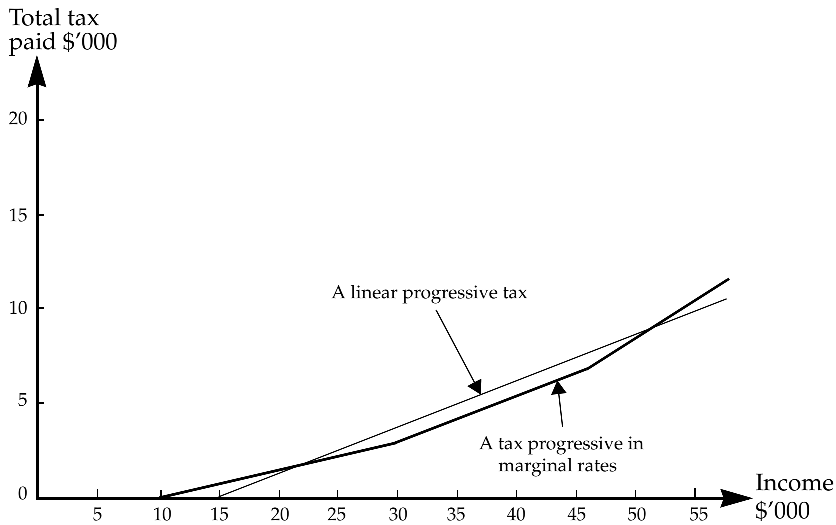

FIGURE 6.2: Taxes under the two tax scales

The critical features of Figure 6.2 are:

1. Taxable incomes between $10,000 and $15,000 have lower average and marginal tax rates under the linear progressive tax.
2. Taxable incomes between $15,000 and $22,500 have a lower average, but a higher marginal tax rate under the linear progressive tax.
3. The marginal and average tax rates are lower under the original tax scale for incomes between $22,500 and $30,000.
4. For taxable incomes between $30,000 and $46,000 average tax rates under the original tax scale are lower, while marginal rates under the two scales are identical.
5. Taxable incomes between $46,000 and $51,000 have a lower average, but a higher marginal, tax rate under the original tax scale.
6. Taxable incomes above $51,000 attract lower marginal and average tax rates under the linear progressive tax.

Assume, to begin with, that incentives to earn income and avoid and evade taxes are unaffected by the change in tax scales. The change in after-tax incomes brought about by the change in tax scales could then be arranged to look like a move from distribution A to distribution B in Figure 5.9 on page 136. Both the lowest- and the highest-income groups are made better off, while the middle-income groups are made worse off. By choosing the parameters appropriately, this change can be arranged so that the mean and variance of the after-tax income distributions are unaffected (assuming the before-tax incomes remain unchanged). A person who cares about negative skewness in the income distribution should favour replacing the original progressive tax scale by the linear progressive one.

In addition, however, before-tax incomes are likely to rise when the original scale is replaced by the linear one. The reduction in average tax rates on incomes above $51,000 would greatly reduce the incentive for these people to hide income through manipulating accounts and so forth. The slightly higher average tax rates on incomes between $22,500 and $52,000 would tend to work in the opposite direction. The differences in average tax rates are slight, however, while these people are less likely to be paying sufficient tax to make such schemes worthwhile.

The lower average tax rates under the linear scale for incomes below $22,500 may also provide strong incentives to work for people whose attachment to the labour market is weakest. Again the slightly higher average tax rates on incomes between $22,500 and $52,000 are unlikely to discourage many of these people from participating in the labour market at all.

Finally, the higher marginal tax rates under the linear scale on incomes between $15,000 and $30,000 are likely to provide a greater disincentive for increasing income at the margin. This is likely to be more than offset, however, by the substantially lower marginal tax rates on incomes between $10,000 and $15,000 and all incomes above $46,000. Many people in the latter group, in particular, may have substantial opportunities to earn additional pre-tax income.

In summary, a change from a progressive to a linear tax intended to preserve the mean and variance of post-tax incomes is likely to raise pre-tax income. Tax collected would most likely rise, allowing a further reduction in the marginal tax rate. The efficiency losses caused by the tax, which are likely to depend on the square of the tax rate, are thus likely to be much lower under the linear tax.

What therefore could explain opposition to a flat tax on the grounds that it is inequitable relative to an income tax that is progressive in marginal rates? Substituting a flat tax for a progressive tax will raise the after-tax income of the poor relative to the middle class, and should on that account be regarded as an 'equitable' change. For the change to be inequitable, the concomitant increase in after-tax income of high-income earners relative to the middle class must be regarded with substantial hostility. This hostility also has to be sufficiently great to offset the benefits of lower taxes overall, including lower efficiency losses.

People who dislike replacing a tax that is progressive in the marginal rates by an equal-revenue flat tax with a higher zero bracket amount must value negative skewness and dislike positive skewness. Regardless of their rhetoric, they must be motivated by envy and not by compassion.

The most likely explanation for the political appeal of an income tax progressive in marginal rates is that it tends to tax middle incomes more lightly than a linear tax that raises the same revenue. The marginal voter needed to obtain a majority is, however, likely to be of middle income. Blum and Kalven (1978: 19) note that there may be a danger in this process:

> Under any progressive system today the higher surtax rates are almost certain to apply only to a minority of voters. This means that a majority are allowed to set the rates which fall exclusively on the minority … A proportionate tax on income presents no comparable weakness since each voter in voting taxes for others to pay necessarily votes an equivalent rate for himself.

Progressive taxes versus welfare payments

We have argued that a tax that is progressive in marginal rates is worse than a linear tax with a higher exemption level. We do not mean to imply, however, that high-income exemption levels come at no cost. On the contrary, allowing any income to be exempt from tax imposes many costs.

If the government wishes to raise significant amounts of revenue, most taxpayers need to face a positive marginal tax rate. The major group of people with incomes below the exemption level would be part-time workers, or people such as new entrants to the work force or retirees who worked for only part of a year. It may include spouses who earn money only a few hours each week, students who work during the holidays, people engaged in income splitting, such as children in receipt of interest and dividends, and self-employed people who have a bad year.

For those facing a positive marginal rate, the exemption level represents a loss of tax revenue for the government, which has to be made up (for the same total level of tax receipts) by a higher marginal tax rate. The higher marginal rate would impose efficiency losses, and since the overall revenue raised has to be the same, there would be few if any efficiency gains from the changes in average tax rates for different people. In particular, the efficiency losses from the higher tax rate on almost all taxpayers would more than offset the efficiency gains from a lower tax rate on the people with very low earnings. This likelihood is compounded by the fact that the labour supply and other choices of many people who have incomes below the exemption level may not be very sensitive to the tax rate. Decisions about when to join or leave the labour force, or whether to earn income in school holidays, are likely to depend on many factors other than after-tax income in one year. These decisions are often related to lifetime plans or long-term investments.

As we noted above, allowing an income exemption incurs many of the administrative burdens of a progressive tax. Withholding taxes on interest and dividends, and the taxation of income earned through companies, trusts and superannuation funds become more costly to operate than under a flat tax.

Finally, an income exemption level represents a very poorly targeted form of assistance to the poor. As we noted, many of the people benefiting from an income exemption for tax purposes will not be poor, particularly when we consider their expected lifetime incomes and their access to intra-family transfers, rather than their market incomes in any one year.

A far better system involves taxing everyone at the same tax rate and using the welfare system to provide transfers to the poor. We would need the administrative apparatus to provide a welfare system even if the tax has an income exemption, since a zero rate of tax on a small amount of market income does nothing to assist those who are genuinely poor and earn no market income. The welfare system is also perhaps better suited to examining the detailed situation of each potential recipient and taking into account special factors such as family circumstances to determine when support is needed.

Income tax rates in New Zealand

Consistent with the arguments presented here, many countries, including New Zealand, have in recent decades reduced the number of different marginal rates in their income tax system. None of these countries has been bold enough to capture the full benefits from moving to a completely flat rate structure. Some politicians in the United States have, however, recently declared themselves in favour of such a change.

Table 6.3, reproduced from Statistics New Zealand (1999: 16), shows the rates applicable in New Zealand in 1982, 1986 and 1996. The figures relate to income years for taxpayers in New Zealand which end on 31 March. The rates of income tax which will apply from 1 April 2000 are set out in Table 6.4. A new top rate of tax of 39% on income over $60,000 a year is to be introduced. From 1 July 1998 until 31 March 2000 the tax scale was the same as that presented in Table 6.4 except that all income over $38,000 was taxed at the 33% rate. It should be borne in mind, however, that the low-income rebate does not apply to investment income.

The difference between a 15% marginal rate and a 33% or 39% rate is smaller than the difference between a 14.5% rate and a 55% or 60% rate in 1982. When

TABLE 6.3: Historical New Zealand income tax rates

1982		*1986*		*1996*[a]	
Income (dollars)	*Tax rate (percent)*	*Income (dollars)*	*Tax rate (percent)*	*Income (dollars)*	*Tax rate (percent)*
Up to 5,500	14.5	Up to 6,000	20.0	Up to 30,875	24.0
5,501–12,600	35.0	6,001–25,000	33.0	Over 30,875	33.0
12,601–17,600	48.0	25,001–30,000	45.1		
17,601–22,000	55.0	30,001–38,000	56.1	Low income rebate	
Over 22,000	60.0	Over 38,000	66.0	Below 9,500	15.0

[a] The low income rebate creates an effective three-rate scale, with the effective rate from \$0 to \$9,500 being 15%, and from \$9,500 to \$30,875 being 28%.

TABLE 6.4: Current New Zealand income tax rates

Before rebate		*After rebate*	
Income (dollars)	*Tax rate (percent)*	*Income (dollars)*	*Tax rate (percent)*
Up to 38,000	19.5	Up to 9,500	15.0
38,001–60,000	33.0	9,501–38,000	21.0
Over 60,000	39.0	38,001–60,000	33.0
		Over 60,000	39.0

comparing the rates in Table 6.3 and Table 6.4 we should also note, however, that the rates for 1996 and beyond understate the true tax burden relative to previous periods because a goods and services tax (GST) of 12.5% is also paid now on income when it is spent. Nevertheless, in 1996 New Zealand had a system closer to a flat rate of tax than most countries, and recent changes have moved the system in the opposite direction. This might appeal to some marginal voters, but it is not good policy. As just one example of the way in which making the tax scales more progressive complicates the system, the government is now looking at ways of addressing the over-taxation of low- and middle-income earners who invest in superannuation funds that are subject to a tax rate of 33%. Any solution other than a return to a flatter tax scale will add complexity to the system.

COMBINING TAXES AND TRANSFERS

An important issue that needs to be considered when income transfers are means-tested is that the rate of abatement of benefits adds to the tax rate in imposing disincentives for people to work. Each dollar of additional market income leads to an income tax liability, a consumption tax liability (assuming the after-tax income is eventually all spent) and a loss of welfare benefits (as a result of the means test). Cox (1998, Chapter 4) discusses some of the issues involved in choosing abatement rates for benefits.

Economists used to argue that benefit abatement rates (as a result of means-testing) ought to be kept low to minimise overall disincentives for beneficiaries to work. More recently (see for example Blinder and Rosen (1985)), this view has been challenged by an alternative approach based on the notion that benefits ought to be withdrawn at a high rate over a very narrow income range. People would have little or no incentive to earn additional income within that range, but a modest increase in income could take them completely outside the range over which means-testing applies. People receiving benefits can have a reasonable hope that a modest improvement in their circumstances will free them from reliance on welfare. By contrast, a programme that abates benefits gradually over a wide income range leaves many people in receipt of welfare. Administrative costs are raised, and recipients may have little hope that they can restore their dignity by becoming self-reliant.

Some other features of the combined tax and welfare system that we are advocating may reduce the problem of high effective marginal tax rates for beneficiaries. Once transfers, and programmes such as education and health care subsidies, have been restricted to the poor, overall government spending, and the tax rates required to support it, will be much lower.

Replacing the progressive tax by a flat tax may also allow average, as well as marginal, rates of taxation to be reduced. In particular, a flat tax with no income exemption, and a higher tax rate on infra-marginal income, would raise the average taxes paid by many high-income taxpayers even while lowering their marginal tax rates. The higher revenue raised from higher-income taxpayers would therefore be accompanied by a lessening of the disincentive to earn or report marginal income.

Finally, the efficiency costs of high abatement rates for recipients have to be balanced against the efficiency costs of higher tax rates on all taxpayers. When benefits are phased out more gradually, the total level of government expenditure and the tax rates required to pay for it will both be higher. A high effective marginal tax rate for a small number of people might impose lower efficiency costs than a marginally higher effective marginal tax rate on all middle- and upper-income taxpayers.

Ideal abatement rates might also differ across types of welfare programmes. In cases where incentives for the recipient to become self-reliant are less important, such as invalid pensions, abatement rates could be much lower. For some types of benefits, objective evidence is available concerning the degree to which the recipient suffers a particular disadvantage. Such evidence may enable benefits to be more generous without precipitating government revenue losses due to moral hazard or adverse selection problems. Indeed, some types of benefits, such as sickness insurance, could be offered by private insurance companies, as they have been in the past. Even when the government provides benefits that could be supplied by private insurers, the government could reduce moral hazard by requiring reviews of objective evidence by a third party of its choosing, rather than one chosen by the recipient. Moral hazard can also be controlled, for example, by rules regarding how long a person can remain on benefits without moving to a new location where employment prospects may be better.

CONCLUDING REMARKS

The general problem (noted in Chapter 3) of judging politicians by their intentions, rather than on what they achieve, appears particularly severe in regard to income transfer policies. Spending money is considered equivalent to solving the problems. It appears as if voters want the income transfers, not what the transfers might achieve.

Few commentators question the extent to which income transfers benefit the ostensible recipients. In particular, very little attention is focused on the less obvious or more remote consequences of income redistribution policies. For example, government welfare programmes have displaced intra-family transfers and assisted, if not encouraged, family breakdown. They have induced people to depart from sensible lifetime savings plans, or prudent insurance arrangements. Transfers have also discouraged many recipients from seeking labour market experience, with adverse consequences for their future earning prospects, their self-esteem and the likely future prospects for their children.

Most spending justified on the grounds of equity, such as public spending on education and health, accrues to people who are not poor. These people would benefit more from reduced taxes, for they could then buy higher quality services at a lower price in a competitive market place.

Discussion of taxation also rarely deals with the substantive issues. It is clouded by the fanciful assumption that taxes have no effect on incentives or individual behaviour. Taxation is presented as involving the redistribution of a pile of money the size of which is fixed even though both earned incomes and reported taxable incomes depend on the tax regime.

The key issue in tax policy is the extent to which a compassionate concern for the poor ought to be balanced against the efficiency losses of high tax rates. A flat tax accompanied by cash assistance targeted at the poor, or possibly means-tested vouchers for education, health care and other services deemed necessary, would maximise the benefits from assistance per dollar of cost. Taxes that are progressive in the marginal rates focus attention instead on redistributing income to the marginal middle-class voter who is not poor or deserving of compassion, and may often have a higher standard of living than many of the intended targets of high marginal tax rates.

APPENDIX: SACRIFICE THEORY OF TAXATION

The theory begins with the premise that monetary incomes correlate perfectly with standards of living. Specifically, the utility of person i is assumed to be a function of the after-tax money income of i (see the related discussion in Chapter 5). If y_i is the money income of i, and the average tax rate on that income is τ_i, the after-tax income will be $y_i - y_i\tau_i$ and utility will be $U(y_i - y_i\tau_i)$. The 'equal sacrifice' theory then postulates that tax rates τ_i should be chosen so that the required revenue is raised, while imposing the same loss in utility on each person i. Figure 6.3 illustrates the choice of taxes for just two people in the economy and when the revenue to be raised is R.

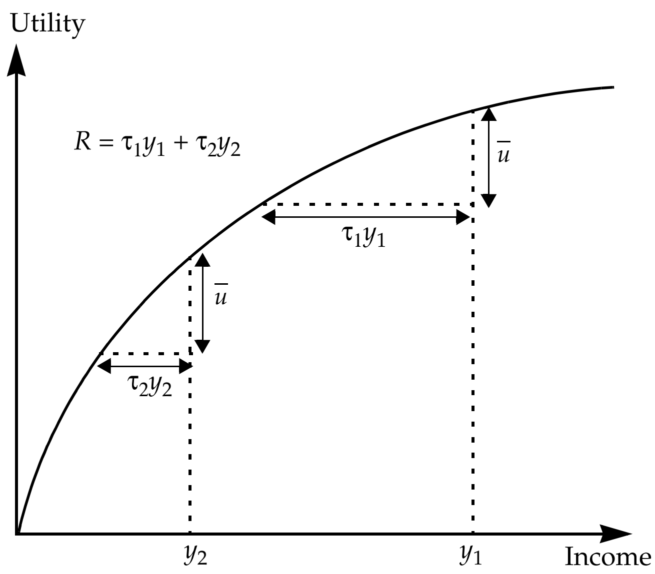

FIGURE 6.3: Income taxes under the equal sacrifice principle

A clear implication of Figure 6.3 is that total taxes (measured by the horizontal arrows) should increase with income as long as the marginal utility of income declines. It does not follow, however, that the tax rate should increase with income. For example, if utility equalled the logarithm of income:

$$U(y) = \log(y) \quad \text{(EQ 6.1)}$$

the marginal utility of income would be:

$$U'(y) = \frac{1}{y} \quad \text{(EQ 6.2)}$$

and would thus decline with income as the argument underlying Figure 6.3 assumes. Yet, since the difference in the logarithm of two numbers equals the logarithm of the ratio of the two numbers by the definition of a logarithm, the loss in utility for any tax rate τ would be

$$U(y) - U[y(1-\tau)] = \log(y) - \log[y(1-\tau)] = \log\left(\frac{1}{1-\tau}\right). \quad \text{(EQ 6.3)}$$

Equal sacrifice would require the same loss in utility for each person and thus, in particular, a loss of utility that is independent of income. Equation (6.3) then would imply that the tax rate should be a constant regardless of the level of pre-tax income y.

The utility of income could, however, be less concave than the logarithm case of equation (6.1). For example, suppose

$$U(y) = \sqrt{y}. \quad \text{(EQ 6.4)}$$

Utility of income curves like equation (6.4) also display decreasing marginal utility as income increases. In place of equation (6.3) we would have:

$$U(y) - U[y(1-\tau)] = \sqrt{y} - \sqrt{y(1-\tau)} = \sqrt{y}[1 - \sqrt{(1-\tau)}]. \quad \text{(EQ 6.5)}$$

If the sacrifice in equation (6.5) is constant at, for example k, then the equal sacrifice theory would require tax rates:

$$\tau = 1 - \left[1 - \frac{k}{\sqrt{y}}\right]^2 \quad \text{(EQ 6.6)}$$

which decrease as y increases.

It might appear that an egalitarian could remedy this unfortunate situation by proposing that we share the burden to keep the proportionate sacrifice constant. In that case, tax rates would need to be chosen as a function of y to ensure $U[(1-\tau(y))]/U(y)$ is constant. As noted by Blum and Kalven (1978: 43):

> The proportionate sacrifice standard would clearly result in progression with any utility curve that would satisfy the requirements for progression under equal sacrifice; and obviously it would also do so under some less steep curves as well. Indeed it would seem to do so for any money utility curve which declines at all. This appears to follow from the observation that if money had a constant utility, the proportionate sacrifice standard would result in a proportionate tax. To impose the same percentage of sacrifice on all taxpayers would then simply require taking the same percentage of dollars from each of them. Therefore, it might seem that, no matter how gentle the slope of a declining utility curve, it

> would be sufficient under the proportionate sacrifice standard to produce something more than a proportionate tax and thus to produce some degree of progression. Over sixty years ago, however, the error of this reasoning was discovered by a Dutch economist, Cohen-Stuart, who elaborately analysed the consequences under the proportionate sacrifice standard of various utility curves for money. He demonstrated that it is possible to construct utility curves which declined in such a fashion that, on some parts of the curves, a regressive tax followed from application of the proportionate sacrifice standard.

As we observed in Chapter 5, if the objective is to minimise the total sacrifice, all people paying taxes need to be left with the same marginal utility of money. With everyone having the same utility of money income curve, equal marginal utility of income will require

$$U'[y_i(1-\tau_i)] = U'[y_j(1-\tau_j)] \qquad \text{(EQ 6.7)}$$

so that after-tax incomes must be equal:

$$y_i(1-\tau_i) = y_j(1-\tau_j)\,. \qquad \text{(EQ 6.8)}$$

Anyone earning more than the average income of all taxpayers would face a marginal tax rate of 100%, at which point one is led to ask why the government would raise any revenue at all. As we argued in this chapter, the sacrifice theory needs to be augmented by an analysis of the incentive and efficiency consequences of taxes in order to arrive at a satisfactory analysis of the costs of taxation.

7

OTHER NOTIONS OF SOCIAL EQUITY

EQUALITY OF OPPORTUNITY

The expression 'equality of opportunity' is used in at least two different ways. On the one hand, equality of opportunity can mean something similar to fair play, or the application of a set of rules equally to all. An analogy can be made to the 'level playing field' idealised in sport. Players should have the same opportunities to attain their goals, although people in similar circumstances may choose to do different things. Players have different chances to succeed because they differ in abilities and other endowments, but the endowments are accepted as an integral part of each person's identity. Fairness extends only to allowing people to do the best they can with their endowments. There should be no artificial institutional or legal obstacles to individual achievement. The concept of fair play is similar to what we have been calling horizontal equity and equality before the law.

A quite different notion of equality of opportunity is closer to equality of outcome. Those who support equality of outcome deny that people ought to have the same opportunity to exploit their endowments. Rather, supporters of equality of outcome claim that inequalities in endowments ought to be addressed in an attempt to ensure that each person has an equal chance of ending up in any given situation. Instead of a level playing field, people ought to be faced with a 'tilted playing field' with hills placed in front of those who are better endowed at the start of life.

The second notion of equality of opportunity nevertheless differs from equality of outcome in several ways. Advocates of this notion of equality of opportunity recognise that even if two people have the same chance to succeed in life, they may not experience the same outcomes. On some occasions one outcome will occur, while on other occasions one of the other possibilities will eventuate. Furthermore, no two people have the same preferences. Even if two people have the same endowments and experience the same sequence of random events, they are unlikely to end up in the same circumstances. It may be much harder, therefore, to ensure that everyone ends up equal than it is to ensure that everyone has the same chance of ending up in any given circumstance. Redistributing to ensure equal opportunities might be thought a more achievable goal than redistributing to achieve equal outcomes. We shall argue, however, that this impression is mistaken.

Measuring opportunities

Equality of opportunity may seem appealing because opportunities are more objective than utility or welfare. We argued in Chapter 5 that utility is a subjective concept. By contrast, opportunities that could be exploited to increase utility are in principle measurable. We could measure the total time available to each person, and observe the possible wages the person could earn by using some time to produce goods or services that have market value. We could also look at the potential non-market uses of a person's time. These could include things done for personal benefit or things that the person does for others on a voluntary basis, although objective information on these is not likely to be available to anyone other than the person involved.

A subset of the opportunities available to people are the possible market goods and services they can consume, or what economists refer to as their budget sets. These types of opportunities are more amenable to measurement than are opportunities more broadly construed. It might be thought that data on market prices and lifetime expected income or wealth would suffice. Unfortunately, many components of wealth are difficult to measure accurately, not least of which is the income-earning capacity of an individual's skills, ability and knowledge. It may also be difficult to know exactly which market goods and services are accessible to people, and what other factors, such as limited time or transport facilities, may have constrained their consumption opportunities. Furthermore, in order to measure lifetime consumption opportunities, we would need data on available interest rates and other investment or insurance options that enable resources to be transferred between time periods, or between a situation in which an insurable event has occurred and one in which it has not.

Suppose, however, that we could show that the choices available to one person *B* are a proper subset of the choices available to another person *A*, so that *A* could choose any of the opportunities available to *B* but not vice versa. We could then state unambiguously that *A* has more opportunities than *B*. Opportunities appear more amenable to objective verification than the utility or welfare that people obtain from exploiting those opportunities as they see fit.

In most cases, we would be likely to find, however, that the opportunities available to one person are neither a subset of, nor do they contain, the opportunities available to another person. Even identical twins raised in the same household are likely to face sets of opportunities that are not comparable in the way we have described.

In addition, while opportunities are in principle measurable, in practice we are at best likely to know only a subset of the real options that another person has confronted. To know all the opportunities available to another person one would need to know all the possible choices that person could have made, not just the choices that were in fact made. Furthermore, these choices pertain not only to the disposition of the individual's money income, but also to potential uses of other available resources such as time or services offered by parents or other family members. Even the total amount of time available for various uses will differ across people. It is not just that one's lifespan depends on the occurrence of many unforeseen random events. Even the probability distribution of potential lifespans will differ across people. People also differ, for example, in the amount of sleep that they need, or the amount of energy they can exert, and this can make an enormous difference as to what can be achieved in a lifetime.

Equality at the start of life

Another notion of equality of opportunity concerns the advantages with which a person is initially endowed, rather than the opportunities that a person faces throughout a life. On this view, if endowments are equalised, the outcome will also be fair, even though random events mean that the final outcomes will be unequal. The pursuit of equality of initial endowments is defective for many of the same reasons that desire for equality of outcome is flawed.

It is not easy to discern which type of initial endowments will be valuable in a person's lifespan. An ability to invent a great computer programme might have been realised in some other marketable skill less than half a century ago, but that skill may not have had as high a payoff as the same skill has today. A great ability to play basketball or some other popular sport did not have the same value before the invention of television. A superior skill in controlling draught horses became much less valuable after the invention of the internal combustion engine, and so on.

Since opportunities cannot in practice be measured any better than can welfare, money income or wealth again becomes the focus of attention. Inheritance of financial or physical wealth is considered unfair, although no objection is raised to other inheritances that might prove far more valuable. Equalising inheritance of financial or physical wealth alone might, however, exacerbate inequalities of opportunity more broadly construed.

Bequests of money or property may also have a very uncertain future value. For some people, bequests may even end up being of negative value. A recipient encouraged to live a profligate life may subsequently suffer an unfortunate accident and be left with few resources or marketable skills. Many wealthy people who are aware of this possibility require their children to attain a certain age before receiving distributions from a trust. The parents hope that their children will be forced to invest in skills and good habits before they receive a large amount of money. Unfortunately, the parents' desires are often not fulfilled by their children.

An objection to unequal inheritances cannot be motivated by a concern that people enjoy at least a minimum standard of living. As with concerns over unequal outcomes, a concern for unequal inheritances only makes sense if people are envious of those who appear to be better off. It does not follow from a compassionate concern for people in dire circumstances. The latter may imply that support ought to be offered to people with particularly disadvantageous endowments, but it would not imply that endowments ought to be reduced in order to achieve greater equality. Indeed, a concern over unequal inheritances of financial and physical wealth is more problematic in this regard.

A compassionate concern for people living in poverty is unavoidably related to outcomes. It could not justify a particular concern with low endowments rather than any other cause of poverty. For example, recent immigrants to New Zealand, Australia, the United States and other similar countries begin their lives in their new country in a situation of unequal opportunity. They often do very well, however, in spite of their handicaps. It is not compassionate to transfer resources to these people merely because they started out as unequally endowed.

Compassion motivates us to help people who are in serious difficulties regardless of the reason they happen to be in such a state. There is no good reason to focus on the resources that destitute people did or did not inherit. Inheritance is just one factor among the millions that help determine a person's situation in life.

Taxing inheritances also imposes efficiency losses analogous to those associated with redistributing income to reduce unequal outcomes. Some parents are motivated by a desire to have their children succeed. If parents are frustrated in their desire, they will have less of an incentive to earn income in the first place. Redistribution is not a zero sum game. The value of a transfer to the recipient is substantially less than the loss in benefits to the taxpayer. We are also led to ask why a desire to spend on one's children ought to be considered less worthy than a desire to spend on oneself:

> Instead of being recognised as essential institutions for raising responsible citizens through moral training and developing character, families are seen as the source of unfair differences. (Green, 1996: 51)

We can see one situation, however, in which a notion resembling equality of opportunity may have merit. Normally, the respect for individual autonomy extends to allowing parents to decide how to use family resources. A family in desperate need is deserving of material support. When resources are offered, however, the parents may continue to neglect their children. In order to provide opportunities for such children, we may need to override the wishes of the parents. Redistribution in the form of vouchers for educational or health services that could only be used by children might ensure that the children in question receive essential assistance while leaving substantial scope for parental autonomy.

Income mobility

In a free society, people in positions of privilege continually face competition from other, more able, and ambitious challengers. Anecdotal evidence suggests that families can move from low-income levels to high-income levels and back again in a few generations. Although it is difficult to obtain systematic evidence of changes in income over time and among generations, there is a general presumption that freedom and competition tend to promote greater income mobility. Freedom preserves opportunities for the disadvantaged of today to become the privileged of tomorrow. In struggling to improve their lot, however, the disadvantaged of today help others by doing what people are most willing to pay them to do.

Income mobility in the United States

The best data on income mobility can be found in the longitudinal data sets that record information about income (and many other variables) for the same sample of individuals over many years. These are available only in the United States. The data covering the longest period is the Panel Survey on Income Dynamics (PSID), carried out by the University of Michigan. Cox and Alm (1999: 72–78) discuss some revealing statistics arising from their analysis of a sample drawn from this data set. The statistics are reproduced in Table 7.1.

The procedure the authors used to derive the numbers in the table can be described as follows. First, they obtained, from official Census Bureau statistics, the income quintiles for the part of the US population considered to be 'active' members of the labour force in the years 1975 and 1991. That

TABLE 7.1: Movement between income quintiles, 1975–91

	Percent in each quintile, 1991				
Income quintile, 1975	*1st*	*2nd*	*3rd*	*4th*	*5th*
1st	5.1	14.6	21.0	30.3	29.0
2nd	4.2	23.5	20.3	25.2	26.8
3rd	3.3	19.3	28.3	30.1	19.0
4th	1.9	9.3	18.8	32.6	37.4
5th	0.9	2.8	10.2	23.6	62.5

includes the employed, those laid off, the unemployed, students and retirees, but excludes adults choosing to work at home, youths, prisoners, military personnel, the permanently disabled and the mentally ill. The 20% of the labour force with the lowest incomes in a given year constitute the first income quintile. The second quintile consists of those earning more than at least 20% of all members of the labour force, but less than at least 60%. The third quintile (or middle 20%) contains those earning more than at least 40% of all members of the labour force but less than at least 40% of all labour force members. Similarly, the fourth and fifth quintiles are the remaining two groups of 20% after active labour force members have been ranked by their income levels. In accordance with standard US government definitions, income is defined to include "wages, investment earnings, pensions and government transfers such as social security, unemployment benefits and welfare" (Cox and Alm (1995)).

The income of each individual in the PSID sample in 1975 is then categorised relative to the income of the US active labour force in that same year. The authors found that the income distribution in their sample in 1975 approximated the US personal income distribution among labour force members in the same year.

The first row of Table 7.1 refers to all individuals in the sample who earned an income that would place them in the lowest 20% of earners in 1975. Similarly, the other rows refer to the households in the sample that fell into each of the corresponding income quintiles for 1975.

The columns in Table 7.1 then examine where these people ended up in 1991 relative to the US income distribution among labour force members in 1991. Thus, for example, of all the individuals who started out in 1975 with incomes that would have placed them in the bottom 20% of all members of the US labour force, only a little over 5% of them ended up in 1991 with incomes that would have placed them in the bottom 20% of all members of the US labour force in 1991. Eighty percent of the people starting in this lowest quintile in 1975 ended in the highest three quintiles – 'middle income or better' – by 1991. Almost 30% ended with an income in 1991 that would have placed them in the top 20% of household incomes in 1991.

Less marked upward mobility characterises other individuals starting further up the income distribution in 1975. Almost 24% who started out in the top quintile of the distribution in 1975 had moved down to the fourth quintile by 1991, while another 10% had fallen to the third quintile.

Cox and Alm also found that less than 1% of the individuals in their sample remained in the bottom quintile of the corresponding US income distribution for every year of their sample:

> More than half of those in the lowest 20 percent in 1975 had reached one of the top three tiers within four years. Two-thirds of these people made that leap within six years, and three-fourths did it in nine years. (Cox and Alm, 1999: 74)

The statistics underlying Table 7.1 relate to individual and not family income. Some of the people with low income in 1975 are undoubtedly young people just starting their careers and some may only have worked for part of the year in 1975. The conclusion that such people move up the earnings distribution as they age is, however, an important part of the income mobility story.

Cox and Alm cite other data from the US Department of Labor reporting the median length of a spell below the official poverty line (the median length is that length that has an equal number of lengths greater and less than it). In the early 1990s, the median length of such a spell was 4.2 months. At that time, only one-third of the people classified as being below the poverty line by the Census Bureau had been there for 24 months or more.

A noticeable feature of Table 7.1 is that the tendency to move to higher quintiles is greater than the tendency to move back down to lower quintiles. If this happened systematically, the lower-income quintiles would become depopulated. By definition, however, 20% of the labour force has to be in the bottom quintile. The explanation is that while the sample is representative of the entire US labour force in 1975, it is no longer representative in 1991. Recall that people in the sample were tracked continuously for 17 years. By 1991, these people had accumulated considerable labour market experience, which is in turn reflected in their higher incomes. The places of those who came from the lowest quintile in 1975 would be filled in 1991 by new entrants to the labour force beginning their own climb up the employment ladder. The results reported by Cox and Alm also reflect the lifecycle hypothesis of earnings growth which we discussed in Chapter 5.

Cox and Alm report some other related results from the PSID data. They also discuss a US Treasury Department study that was based on nine years of taxation data:

> … researchers found that 86 percent of those in the lowest 20 percent of income earners in 1979 had moved to a higher grouping by 1988. Moreover, 66 percent reached the middle tier or above, with almost 15 percent making it all the way to the top fifth of income earners. Among Americans who started out above the bottom fifth in 1979, the Treasury found the same movement up the income ladder. Nearly 50 percent of those in the middle tier, for example, rose into the top two groupings, overwhelming whatever downward mobility that took place. (Cox and Alm, 1999: 76–77)

The evidence of income mobility from these studies can be contrasted with two other types of evidence. Cox and Alm also report that US Census data reveal that the highest quintile of households, measured by total household (not individual) income, have been earning an increasing share of national household income at least since the early 1970s. (Actually, the share of the highest quintile tends to decline in recessions, but the trend has been quite positive.) Conversely, Cox and Alm (1999: 71) observe that the share of household income garnered by the lowest quintile of income earners declined from 4.4% in 1975 to 3.6% in 1997. Such changes in income share tell us little about whether inequality is increasing or decreasing, however, since the individuals found in each quintile tend to change over time. The poorest quintile is not a fixed group of people but rather a changing group of people as those who were poor improve their situations, climb to higher-income quintiles, and are replaced by new entrants at the bottom of the income ladder.

There has been much debate within the labour economics literature in recent years concerning the increase in inequality in wage rates in the United States. Care should be taken in translating wage rate inequality into inequality in lifetime income, since the same individuals are not destined to remain in any one wage category. It is possible for higher wage differentials to reflect greater opportunities for people to change their incomes by investing in those factors being rewarded more highly in the market place. Furthermore, labour incomes depend on hours worked as well as wage rates, while most households rely on more than one labour income and also earn substantial non-labour income. Nevertheless, some commentators have expressed concern that the increased dispersion in wage rates will result in increasing income inequality.

Welch (1999) produces an interesting overview of the literature concerning inequality in wage rates, along with an analysis of 30 years of data from the US Current Population Survey (CPS). Unlike the PSID, the CPS does not track the same individuals continuously. Rather it is a revolving sample. Up to one-half of the people who are in the sample in any given year will also be in the sample the previous year. Similarly, up to one-half of the people in the sample in a given year will also appear in the following year. Welch uses this overlapping sample structure to show that there is “classic regression to the mean” in the CPS. People below average in the current year tended to be higher up the distribution in the previous year or the following year and, conversely, people with above-average income in the current year tended to be lower down the distribution in either the preceding or following year.

Welch then uses the CPS data to investigate the likely source of the increase in wage inequality in the United States in recent decades. He concludes (Welch, 1999: 12) that the growth in wage inequality has resulted from an increased demand for education and skills. He observes that some of the increase in demand for educational qualifications may be more apparent than real. With the continuing decline in the quality of school education, labour market entrants now go to college to learn what used to be learned in high school. Nevertheless, there also appears to have been a genuine increase in the demand for better educated workers. The composition of gross national expenditure has shifted toward professional services, in which educated labour is a more significant factor of production, and away from agriculture and heavy industrial manufacturing, in which unskilled labour is more important. Within manufacturing, there has also been a shift toward products requiring a more educated labour force, such as electrical machinery, computers and aircraft. Even within firms, automation has reduced the demand for unskilled labour while simultaneously increasing the demand for educated workers.

The period of increasing skill premiums from 1980–92 was not associated, however, with increasing inequality for groups of workers starting with lower wages. On the contrary, Welch (1999: 7) shows that the distribution of wages of black men, black women and white women all rose relative to the distribution of wages of white men:

> Increasing inequality [as a function of skill] within groups distinguished by race and gender coincided with reduced inequality between the same groups. (Welch, 1999: 16)

Welch (1999: 13) also shows that the increase in the educational premium was followed by a dramatic growth over the period 1981–97 in the proportion of young people obtaining a college degree. Previously, a decline in the educational premium had also been followed by a decline in the proportion of young people obtaining a college degree. Thus, Welch concludes:

> The increased educational wage premium created opportunities for individual improvement in economic status, and today's job-market-entering cohorts are more highly educated than any before. (Welch, 1999: 16)

Welch also finds that the period 1967–97 has witnessed a "growing proportion of the population that either does not work at all or works very little". These people with a weak attachment to the labour force are overwhelmingly individuals with low educational attainments. Welch uses the CPS data to examine the incomes of these individuals. He finds that the average *household* income for households including men who do not work are around the 25th percentile for all households in the United States. Many of the men are found in households in which someone else earns some labour income, but welfare payments of various sorts are also an important source of non-labour income in these households. Welch (1999: 15) asks rhetorically, "To what extent are our safety nets magnets?". He later concludes:

> During this period of increasing wage inequality, we have implemented a web of subsidy and assistance programs designed to buffer poverty that appear to have drawn many into simply not working. I do not argue that the growth in the proportions of prime-aged men who do not work is the exclusive fault of such programs. Falling real wages have created incentives to search for alternatives to work; among others, the anti-poverty programs have provided such alternatives. (Welch, 1999: 16)

Income mobility in New Zealand

Cox (1998: 25) cites a New Zealand study of income mobility that was based on income tax data:

> In a recent paper, Barker (1996) points to the substantial degree of mobility that exists within the New Zealand income distribution. Using income tax data he estimates that 25 percent of taxpayers in the lowest quintile of the income distribution between 1980 and 1987 had moved up the income scale one year later, and 46 percent had moved up seven years later.

Cox (1998: 26) also cites a 1997 study of the duration of benefit receipt undertaken by the Treasury and the Department of Social Welfare in New Zealand:

> This study followed a group of beneficiaries and investigated how long they remained on benefits and the reason why they ceased to receive benefits. The main findings are as follows:
>
> - Nearly 50 percent of New Zealand's working age population have received benefits at some time during the past four and a half years.
> - Many spells on benefit are of short duration. Fifty percent of the beneficiaries in the study leave benefit during the first six months and 70 percent during the first 12 months.
> - Seventy-five percent of the unemployment and sickness beneficiaries, 41 percent of the domestic purposes beneficiaries, 17 percent of the widows beneficiaries and 16 percent of the invalids beneficiaries left benefit during the first year.
> - Finding work was an important reason for leaving unemployment benefit, but much less so for the other benefits.
> - Around 10 percent of spells on benefits are for three years or longer. The study found that 59 percent of the invalids beneficiaries, 46 percent of the widows beneficiaries and 31 percent of the domestic purposes beneficiaries were receiving benefits after three years.

Creedy (1997, Chapter 7) also provides an interesting analysis of intertemporal variations in taxable income using New Zealand tax data for 1991, 1992 and 1993. He splits the data into different age cohorts for males and females. Mobility within each cohort is modelled by three factors. The first is regression toward the mean, or (Creedy, 1997: 94) "a systematic … tendency for those in higher income groups … to receive, on average, lower proportionate increases than those in lower groups". In addition, he allows:

> … for the possibility that individuals move through the income distribution of their contemporaries in a systematic way. For example, there may be a tendency for significant improvements to depend to some extent on previous success. (Creedy, 1997: 95)

Finally, Creedy also allows an individual's income, relative to the cohort (geometric) mean income, to be affected by random shocks in each year. Creedy concludes from the parameter estimates that:

> … in general there is some regression towards the mean, particularly in the younger age groups, and some negative serial correlation, except for older females. Hence those with relatively low incomes receive, on average, relatively

> larger proportionate increases, and there is no systematic tendency for success to breed success or for failure to lead to further failure. (Creedy, 1997: 101)

The negative serial correlation in income that Creedy found in almost all cohorts except females over age 52 was reasonably large in absolute value and statistically significantly different from zero. These results suggest that, in most cohorts, success measured in terms of a high market income in one year tends to be followed by failure the next year and vice versa. Recognising that the data relate to taxable income, however, some of this tendency might reflect a shifting of income from one tax year to the next rather than real changes in people's prospects.

In order to draw conclusions about longer-run processes, Creedy looks at systematic variations in the mobility parameters as a function of age:

> In the case of males … the degree of regression towards the mean and the extent of variability [of new shocks] … falls with age, at least up to the late 50s. However, there appears to be no systematic variation in the value of [the serial negative correlation parameter] … For females, the variation is more complex … (Creedy, 1997: 102)

Creedy does not interpret the parameter estimates he obtains for females (his table 7.7) but graphing the results indicates that the degree of regression towards the mean rises slightly until about age 25, then falls slightly until age 34, then rises substantially after that. (The estimates of the serial correlation parameter reported in that table appear to be in error since when one plots the implied curve it departs considerably from the point estimates of this parameter by age group reported in tables 7.4 and 7.5.)

Creedy (1997: 105) examines the data for males in more detail and finds evidence of "considerable stability in age-income profiles". In other words, the New Zealand evidence he examines appears to be consistent with the lifecycle model of income profiles discussed in Chapter 5.

Finally, Creedy (1997: 116) uses his parameter estimates to simulate what might happen over a lifetime. He concludes that inequality of lifetime income is less than the inequality of income in any one year.

INTERGENERATIONAL EQUITY

Intergenerational equity is related to equality of opportunity, since both notions refer to resource use by different generations of people. The idea behind intergenerational equity is, however, more abstract. The current generation has access to certain material resources that can either be used up or retained for use by future generations. The paradigmatic example is a mineral deposit. By extracting the mineral it may mean that future generations no longer have access to that resource or have to expend greater resources to acquire it. Genetic resources provide another example. If the current generation causes species of plants and animals to become extinct, those species' genes may be lost forever, barring further advances in genetic engineering.

Is the current generation taking the concerns of future generations into account when it makes resource allocation decisions? The New Zealand Ministry for the Environment (1995) claims that:

> Future generations are not 'traders' in the market, yet many consider that they should have fair access to resources … The ability [of future generations] to meet their own needs should not be compromised by decisions taken today.

Similarly, the New Zealand Ministry of Fisheries (1996) claims that "intergenerational equity" requires that resource use should not "jeopardise the prospects of future generations". The claims from the two ministries are too strong, since we should consider the welfare of future generations relative to the welfare of the present generation. Using resources in an irreversible way today that raises the welfare of both present and future generations is good.

Do private decisions that affect future resource availability take account adequately of the potential future costs? While future generations are not traders in today's markets, they do have proxies who trade on their behalf. One group of proxy traders is parents buying resources for their children, grandchildren and other descendants. One of the harmful consequences of taxing bequests more heavily than other items of parental consumption is that parents become less inclined to look after the welfare of their own descendants and thus, indirectly, to save resources for all future generations.

Anonymous traders in the form of speculators also represent the interests of future generations in today's markets. A speculator who considers that the current market price of a resource is too low can buy the resource and hold it in the expectation of making a capital gain. In particular, if an exhaustible resource is expected to become more scarce in the future, its price is likely to rise and speculators will have an incentive to buy stocks of the resource now and conserve it for the future.

It might be thought that speculation could not preserve a resource beyond the lifespan of the speculator. In order to benefit from buying a resource, a speculator need not, however, hold a resource until it becomes scarce. As the exhaustion date moves closer, the market price of the resource will tend to rise as more people become aware of the situation. The initial speculator can sell the resource to the next speculator down the line.

A similar argument applies to corporations. Even a shareholder who intends to hold a stock for only one year can expect to get a price for the stock that reflects other investors' expectations of the company's future prospects. In consequence, managers have to look to the future prospects of the company if they want to maximise current share values.

A marketed resource will be conserved as long as a speculator, or manager of a corporation, thinks its likely gain in value will offset the so-called 'opportunity cost'. An opportunity cost is the loss from having funds locked up in a resource rather than invested in some other asset yielding a dividend or interest. In fact, not just any asset would count as a relevant alternative investment. Investing in an appreciating resource will carry risks just as any other investment does. The expected gain from holding a resource (minus any storage or direct holding costs) has to compensate for those risks. The appropriate opportunity cost is therefore the expected rate of return on another asset of comparable risk.

Some environmentalists have argued that market speculation on the future value of resources is unlikely to yield a satisfactory outcome because markets discount future values too heavily. Market interest rates are said to be too high. The environmentalists claim that we need government intervention to ensure that the interests of future generations are adequately represented in today's resource allocation decisions. Several empirical regularities in the outcomes of market and political processes suggest these environmentalists are misguided.

First, recent fiscal history casts doubt on the notion that governments will be far-sighted and will safeguard the interests of future generations. For several decades after the end of the World War II, Keynesian economists set out to change public opinion and remove the implicit constraint on governments to balance budgets except in time of war. Once the constraint was relaxed, most democratically elected governments ran budget deficits almost every year. The consequent accumulation of government debt represents a burden on future generations who will be taxed to pay interest and to retire debt.

Second, governments elected by majorities tend to ignore minority views and opinions. Such governments tend toward a one-size-fits-all mode of service. The high costs of monitoring employees in the public sector may also contribute to the tendency to ignore minority views. Employees who are hard to monitor can, other things being equal, indulge their own preferences to a greater extent. Employee discretion can be reduced by the employer having workers only undertake clearly specified tasks.

Markets typically give consumers more choice. Thus, in a market situation, a lone speculator with an extreme view about the future value of a resource can act on that view without requiring the consent of anyone else. The asset can be bought and conserved for the future in the hope of making a gain for one's heirs, if not for oneself. In a majority voting system, a person with an extreme view may be dismissed as an eccentric.

Third, the high cost of information in the political arena systematically biases policies in certain directions. Voters tend to focus on intentions rather than outcomes. Politicians focus on the short-term and immediate consequences of a policy rather than longer-term and remoter ones. Politicians also pander to vested interests at the expense of the wider community. Politics is not an ideal mechanism for protecting the interests of future generations. A general community concern for the interests of future generations is likely to be hijacked by vested interests in order to provide benefits for people alive today.

Fourth, decentralised market activity has been successful at saving art works (and many other cultural artefacts) over long periods of time. The evidence that markets discount future values more than governments do is weak.

Markets do, however, undervalue some resources, such as wildlife. The difficulty is not discounting *per se* but rather the undervaluation of wildlife now as well as in the future. The basic problem stems from the absence of property rights to wildlife. An unowned resource is without market value, and thus private individuals will have an inadequate incentive to preserve it. Focusing on rates of time discounting is the wrong diagnosis of the problem. Relying on government to look after future generations is also the wrong solution. These issues are discussed in more detail in Hartley (1997).

EQUALITY BEFORE THE LAW

Equality before the law is a fundamental aspect of equity that is essential to our freedom and prosperity. Increasingly, however, the idea of equality before the law has come under attack, as groups of people have been accorded privileges through anti-discrimination statutes. The aim of anti-discrimination legislation is to prohibit any person from restricting the choices or actions of another simply because the latter belongs to a specified group.

An immediate difficulty is that we are constantly differentiating and discriminating among people for justifiable reasons. Parents favour spending resources on their own children. People do more to help their friends than they would do for strangers. We also differentiate among sellers every time we shop. In these cases, discrimination favours people on the basis of relevant characteristics that distinguish them as individuals. Anti-discrimination legislation seeks to bar discrimination on the grounds of general characteristics that are declared to be irrelevant as a basis for judgment and should generally be ignored. Anti-discrimination legislation can thus be consistent with individualism, if it is restricted to protecting individuals from deliberate exclusion or unfair treatment.

Epstein (1995) criticises anti-discrimination statutes when he advocates greater simplicity in the law. His proposals are based in part on his confidence in the ability of people to make good decisions for themselves. He also notes the beneficial effects on incentives, and on the production and use of information, when we allow people to be autonomous. Epstein's criticism of explicit anti-discrimination law is based on the notion that interaction among strangers is better handled by tort law, while contract law should regulate arrangements made on the basis of consensual agreements between parties.

A case for retaining anti-discrimination law might appeal to the fact that people who have suffered from past discrimination may have limited human and social capital. They may, as a result of their limited capital, be susceptible to specific types of torts or 'unconscionable contracts' that may not be adequately covered by existing tort or contract law. Buchanan (1995) discusses a case in which it is alleged that Australian Aboriginals suffered from this type of discrimination in the purchase of insurance. Perhaps this concern could be handled within the framework Epstein is proposing, however, by introducing new types of torts or unconscionable contracts.

Legal remedies may also be warranted in situations, such as in isolated small towns, where reduced competition leads to a *de facto* monopoly supplier or consumer for some goods or services. Increased competition raises the cost of discriminatory economic decisions and tends to eliminate them. A person basing employment or purchasing decisions on irrelevant criteria is, by definition, forgoing economic benefits. For example, prejudice may result in the employment of a less-qualified individual for the ruling market wage. Prejudice may also cause an employer to forgo an opportunity to attract a qualified individual from another employer who is paying a below-market wage. A buyer who is prejudiced may also pay more simply to avoid purchasing a product from a particular supplier. Competition tends to make such practices unprofitable and to eliminate them from the market place. Laws combating anti-competitive actions may therefore reduce the scope for practising discrimination while simultaneously delivering efficiency gains to the New Zealand community. More generally, as Epstein (1995) argues, actions ought to be illegal because they infringe other basic rights and not because the victim belongs to an identifiable group.

The relationship between inadequate competition and discrimination raises another issue. Government is, unfortunately, often the biggest impediment to competition. Restrictions on foreign trade reduce competition for industries producing goods or services that are traded, or would be so were it not for the protective barriers. Government protection of restrictive employment practices, including occupational licensing and monopoly representation by trade unions, also produces significant barriers to competition. In fact, unions and occupational licensing boards have been notorious for discriminating against minorities in the past. Ceasing all interference with normal competitive processes is probably the most effective action the New Zealand government could take to discourage discriminatory practices. This would have the added benefit of delivering substantial efficiency gains to New Zealand as a whole. By contrast, specific anti-discrimination statutes are likely to come at the expense of efficiency while imposing substantial administrative and legal costs.

Anti-discrimination law is a poor way of compensating for past discrimination. The interpretation and application of anti-discrimination statutes has resulted in affirmative action, or unequal treatment in order to correct for past discrimination against people within the now favoured groups:

> Virtually no one has seriously denied that discrimination and bias have resulted in various inequalities. It is the converse proposition – that discrimination or bias can be inferred from statistical inequalities – which is the reigning *non sequitur* of our times, both intellectually and politically … The implicit assumption is that a more or less even or random representation or reward for performance could be expected, in the absence of institutional or societal policies and practices which disadvantage one group compared to others. Yet there has never been an even or random world, even in matters not controlled by the biases of others. Not only performance differences but also differences in luck and in many other factors wholly disrupt the simple picture of an even, regular, or balanced world. (Sowell, 1999: 62)

While legal privileges can certainly make individuals in a favoured group better off, it is not the best way of doing so. Not all the people benefiting from the policy will have suffered as a result of past discrimination. Furthermore, the assistance provided to people who do benefit is unlikely to be related either to their past losses as a result of discrimination or to their current level of welfare. It does not matter, for example, whether a favoured person is in abject poverty or earning an income that places that person well above the poverty line.

Anti-discrimination policies may also do little to redress the problems caused by past discrimination. For example, past discrimination may have left people with a poor education and few skills. These deficiencies would not be addressed by a policy that requires employers to discriminate in favour of people from a formerly disadvantaged, but now favoured, group regardless of whether applicants have adequate schooling or work skills. If worker productivity declines as a result of the policy, the wages of all workers are likely to decline along with it. Assistance with education or training for the disadvantaged individuals within the favoured group (and similarly disadvantaged individuals who are not in the favoured group) would appear to be a more effective policy.

Sowell (1999: 78–79) also makes a more subtle point. An emphasis on discrimination, and anti-discrimination legislation as the appropriate remedy, may dissuade disadvantaged individuals from taking actions to remedy their situation:

> Where poorer people are lacking in human capital – skills, education, discipline, foresight – one of the sources from which they can acquire these things are more prosperous people who have more of these various forms of human capital. This may happen directly through apprenticeship, advice, or formal tutelage, or it may happen indirectly through observation, reflection, and imitation. However, all these ways of advancing out of poverty can be short-circuited by an ideology of envy that attributes the greater prosperity of others to "exploitation" of people like themselves, to oppression, bias, or unworthy motives such as "greed", racism, and the like. Acquisition of human capital in general seems futile under this conception and acquisition of human capital from exploiters, the greedy, and racists especially distasteful.

Another problem with affirmative action is that there is a mismatch between the timeframe of the policy and the ultimate goal. The policy is meant to be only a temporary measure since future discrimination against people in the protected groups has been outlawed. Yet legal privileges, once instituted, become difficult to eliminate once they have served their purpose.

A further defect of affirmative action is that the costs and other effects of the policy are difficult to determine since they depend upon myriad particular facts and circumstances. By contrast, an income transfer to the target group is more open to public scrutiny and its costs in terms of increased taxes, or forgone expenditures on other items, is more transparent.

Such transfers are thus more likely to be reviewed in order to evaluate their merit.

The political advantages of affirmative action have a further consequence. Sanctioning the use of the government monopoly on coercion to discriminate among people in otherwise similar circumstances leaves little guidance in deciding which groups may become so favoured. Any group that can convince the governing majority to support its claims will be able to gain benefits at the expense of others. Different groups will then waste substantial resources in an attempt to become the favoured party.

Affirmative action amounts to an official sanction of discrimination against people who are not members of protected groups. The fairness of processes is judged by their group-distributional results, not by the extent to which judgments are based on general and irrelevant, rather than individual and relevant, characteristics. Someone who is not in a protected group can, for example, be excluded from employment solely as a result of membership in a non-protected group. The person may be considered qualified or able to do the job as well as, or better than, the person actually hired. A type of reverse discrimination occurs in these cases.

Epstein also points out that government discrimination is far more harmful than private discrimination, because it is imposed universally (Epstein, 1995: 178). Policy towards Aboriginals in Australia provides a good example of the harm caused by government discrimination (Warby, 1997).

Furthermore, governments are not well suited to achieving outcomes for individuals because governments cannot be allowed to discriminate, in the sense of treating people differently when it is appropriate to do so. Government must treat people as members of legally defined classes, not as individuals.

The principle of equality before the law ought not to be abandoned. Before using the legal system to enforce discriminatory treatment, we ought first to abandon policies that are inefficient and that harm people disadvantaged by prior discrimination. For example, public education often forces poor children to attend inferior schools. Minimum wage laws hurt people who have not been able to invest in skills or education. Requiring youths to be paid adult wages eliminates many training opportunities for minorities.

As Epstein argues, deregulation of markets tends disproportionately to help those subject to discrimination or those with limited social or human capital. Deregulation lowers the cost of transactions and makes entry to markets easier. It also breaks down cartel arrangements. For example, in the

United States, deregulation of trucking was more effective at reducing racial discrimination against truck drivers than were anti-discrimination laws (Epstein, 1995: 176). Another example is the fall in Maori unemployment (from 28% in 1991 to 14.5% for the December quarter of 1999) after the passage of the Employment Contracts Act 1991.

In a competitive market economy, there are natural social pressures to rectify the harm caused by discriminatory behaviour. An individualistic approach to anti-discrimination policy reinforces these natural pressures. Such an approach affirms that individuals should be judged on their merits, and it provides the means for individuals to be compensated for discriminatory action through normal courts rather than specialist tribunals. An individualistic approach incorporates the notion that horizontal equity and equality before the law are principles worth reiterating in public. An individualistic approach to discrimination is consistent with the idea that institutional arrangements can unfairly exclude or disadvantage certain groups, but it includes no presumption that there is an appropriate or representative distribution (regardless of individuals' interests or preferences) of groups throughout all aspects of society.

Native land title

In the Australian and New Zealand context, a dubious form of official discrimination is the granting of a special form of land title to indigenous peoples. Native land title is discriminatory because it differs from, and is in practice inferior to, the freehold title granted to non-indigenous people.

Citizens of industrialised democracies take their current prosperity, and the average annual growth in prosperity, for granted. Prosperity is not a random event, but results from the evolution of institutions peculiarly suited to the production and maintenance of wealth (North (1996) and Powelson (1994)).

The economic institutions of industrialised democracies are characterised by lower transactions costs than other modes of exploiting resources. Organised markets facilitate the vast range of exchanges that underlie wealth and prosperity by making exchanges easier and cheaper. The institutions of industrialised democracies also help maintain freedom by empowering individuals, and promote prosperity by directing human effort toward wealth creation through voluntary exchanges rather than toward the use and abuse of coercive political power.

Well-defined, tradeable property rights are the backbone of institutions that characterise industrialised democracies. By clearly delineating who decides how resources can be used, secure private property rights reduce transactions and negotiation costs. When people can use their property as they see fit, and can be confident of gaining returns from investments, they have a greatly increased incentive and capacity to discover more effective uses of resources. Decision-making is put in the hands of those most likely to be knowledgeable about the resource and with a direct incentive to extend and use that knowledge. Nominal ownership of a resource does not provide the basis for wealth. It is the secure right to use a resource as one chooses that provides the basis for productive use of that resource as an asset.

Property rights, including those involving land, did not derive from preordained principles. Rather, such rights evolved from rules codifying successful ways of handling clashes of interests. The general tendency was for property law, and land law in particular, to evolve towards simplicity, with increasing security and increasing discretion in its use.

Simple property laws provided for the widest and strongest congruence of interests in a pluralist society. Freehold title, with a single legal owner able to sell and otherwise make decisions regarding land, represents the clearest outcome of the tendency toward simplicity in land law. The alienability of freehold title allows property to pass easily to those who believe they are able to use it more efficiently. Furthermore, a decision-maker who also owns a resource has incentives to maximise the value generated from that resource. Freehold title, with a single legal holder of rights over land, thus provides an efficient form of ownership.

Property rights that are more uncertain, either in their extent or in their security, provide inferior incentives to invest in finding better uses for resources, particularly over the long term. Insecure ownership effectively raises time discount rates, reducing incentives to invest and lowering rates of capital accumulation and economic growth.

In particular, providing inferior land titles to indigenous peoples makes it much more difficult for these groups to improve their lives through their own efforts. Land with an inalienable title cannot be used as collateral for loans needed to raise productivity through fencing, pasture improvement, animal breeding programmes and so on.

More generally, when buyers and sellers of land have to take account of extended kin obligations, incentives to use resources productively are greatly reduced. Under native title, individual rights and responsibilities are poorly defined and weakly enforced. People have reduced incentives to exert effort or make investments for the future. An increase in the number of parties involved in a transaction raises negotiation and transactions costs, which tends to reduce the number of exchanges. Parties with a minor interest in a transaction can refuse to approve a trade in an attempt to extract more benefits for themselves. Exchanges that permit the involvement of only those people who contribute some benefit maximise the incentive to discover such benefits.

Secure, well-defined private property is an essential element in a free society. Unless citizens have resources over which they exercise control, they are at the mercy of those holding coercive political power. As Hayek (1944: 78) pointed out, a relatively lowly paid official, armed with the appropriate legal discretion, has far more power over our life than the millionaire who lives next door. Secure private property provides protection against the vagaries of those in power. An individual living in communal housing on communal lands is at the mercy of communal officials and politics. Communal living arrangements facilitate coercive actions that would not be tolerated in the wider community.

An individualistic approach to property rights allows for greater diversity in decisions concerning resource use and protection of individuals against group decisions. The individualistic approach also means that the benefits of successful decisions, and the penalties of poor ones, are concentrated rather than dissipated, maximising the incentives for effective use. The incentives are to gain maximum individual benefit for minimum cost in time or effort.

One argument in favour of combating discriminatory behaviour and institutions is that a good society places paramount value on freedom and individual opportunity. Because freedom is so essential, indigenous property should in the first instance be individually based, although a free society would allow people, if they so desired, voluntarily to pool resources in a communal way.

The defects of native title need not, however, accompany joint ownership of land and other resources. Modern market economies already have an institution, the common stock corporation, that allows for efficient joint ownership. The joint owners of a corporation retain identified, marketable shares with well-defined rights and responsibilities. Individuals dissatisfied with management of the joint asset, or who no longer identify with the relevant community, could readily exchange their entitlements for something they value more highly. Title that cannot be exchanged for alternative assets is of little or no value to individuals who disagree with communal decisions.

Some argue that the collectivism of traditional societies (including traditional Maori society) is something particularly precious that ought to be preserved. If such collectivism is applied to resource management, however, it is likely to undermine the material basis of a vibrant Maori culture. As Sowell has recently noted:

> Cultures have consequences. Ignoring those consequences while proclaiming equality as a self-justifying ideal does nothing to benefit the less fortunate, and in fact … widens the gap in the name of narrowing it …
>
> In a world where every society and every civilisation has borrowed heavily from the cultures of other societies and other civilisations, everyone does not have to go back to square one and discover fire and the wheel for himself … Yet the equal-respect "identity" promoters would have each group paint itself into its own little corner, with its own insular culture, thus presenting over all a static tableau of "diversity", rather than the dynamic process of competition on which the progress of the human race has been based for thousands of years. (Sowell, 1999: 75–77)

In particular, granting land and associated resources on a communal and inalienable basis greatly reduces their value as assets to the new owners. Such an approach also compromises the prosperity of all New Zealanders by removing the resources from normal economic influences. The failure of command economies, such as those in the former Soviet Union and Eastern Europe, provides a stark example of the deleterious effects of communal, inalienable title. It seems ill-advised to advocate policies similar to those in command economies as a path to advancement for indigenous peoples.

The legal system and redistribution

In the latter half of the twentieth century, the legal system has come to be viewed as a mechanism for redistributing income. A prevailing assumption is that parties who have suffered due to an accident ought to be able to seek compensation from anyone connected with that accident, no matter how remote the connection is. The primary criterion for awarding damages appears to be the extent to which the costs of awards can be spread across many individuals. Compensatory payments effected through the legal system thus become a type of tax.

The rules of liability that evolved under the common law system embody an attempt "to minimise the sum of the administrative (including error) costs and the costs associated with the creation of poor incentives for individual action" (Epstein, 1995: 32). A corollary of this proposition is that if the rules of liability are altered in order to promote redistribution then the efficiency of the legal system will be compromised.

Under a strict liability regime, a party who can take care to avoid an accident (called the 'injurer') is held liable for any damages incurred should the accident occur. The injurer has an incentive to balance the costs of taking care against the expected benefits of doing so. The latter will be reflected in the expected payments the injurer would have to make in the event an accident occurs.

One problem with the strict liability regime is its implication that an injured party is entitled to litigate every accident which occurs. The resulting administrative costs may be high. The rule of negligence can be seen as a way of limiting the number of court cases. Under a negligence regime, an injurer is liable only if it can be shown that the care taken to avoid an accident, or the damages resulting therefrom, was less than reasonable. One can show (see, for example, Hartley and Brito (1996) and Brito and Hartley (1999)) that if the negligence standard is chosen in an efficient manner, 95% or more of potential injurers will meet the standard, leaving at most 5% of cases to be litigated. Injurers have a strong incentive to meet the standard, even when it is high, since by doing so they are exempt from liability.

The dramatic saving in administrative costs under a negligence regime comes at the expense, however, of reduced incentives to avoid an accident. In particular, if the amount of participation in a potentially dangerous activity can be varied, strict liability leaves the injurer with an incentive to choose the level of participation that balances expected benefits against expected costs. In contrast, a negligence regime in which participation is excluded from the negligence standard will not provide incentives for individuals to balance the benefits and costs of participating in a potentially dangerous activity. The level of participation in an activity is usually beyond the purview of the court, however, since the value of engaging in an activity is likely to include significant factors that are not objectively verifiable. One might expect, therefore, that negligence sometimes minimises Epstein's "sum of the administrative (including error) costs and the costs associated with the creation of poor incentives for individual action" while under other circumstances strict liability does so.

Neither strict liability nor a rule of negligence will provide appropriate incentives when both parties involved in an accident can take care to reduce the probability of it occurring. When one party can recover all, or most, of the damages resulting from accidents, that party will have little incentive to take care. This is less of a problem under an efficient negligence regime since, as we noted above, under such a regime the injurer will rarely be negligent. The incentives for an injured party to take care nevertheless can be increased by providing a defence of contributory negligence. This defence absolves the injurer from liability whenever there was negligence on the part of the victim. In effect, there is a standard of care for potential victims as well as potential injurers.

Brito and Hartley (1999) show that, when the level of participation in an activity is not a factor affecting accident probabilities, standards of care for injurers and injured parties can be chosen to achieve outcomes that are close to the most efficient imaginable. The magnitude of excess costs under the most efficient contributory negligence rules depends mainly on the costs to an injurer of taking precautions (care). In particular, the best contributory negligence rule performed relatively better when injurers had lower average costs of taking care. Contributory negligence also performed relatively better when care taken by one party increases the effectiveness of care taken by the other party (that is, care is 'complementary'). Both these results are explained by the fact that care taken under the best contributory negligence rule deviated from the efficient levels by inducing injurers to supply excessive care on average. This is less costly when injurer costs are lower. It is also less costly when care is complementary because the theoretically most efficient solution (which is unattainable in practice) would require relatively high levels of care from both parties regardless of their relative costs.

Potential victims choose close to the efficient amounts of care to take under the most efficient contributory negligence rules. If all injurers met their legal standard, victims could not expect to recover damages. As a result, victims would have an incentive to trade-off the costs and benefits of taking care, and thus choose the efficient amount of care to take. Since almost all injurers meet their standard of care, an efficient contributory negligence regime gives potential victims close to the right incentives.

A negligence regime with a defence of contributory negligence is also likely to result in low administrative costs. Under the efficient negligence standards, typically considerably less than 5% of injurers would fail to meet their legal standard of care. In addition, there is a discrete gap between the legal standard of care and the care of those injurers who do not meet the standard. Hence, negligent injurers should be easy to identify. Litigation should occur only in the small percentage of accidents in which the injurer has not met the legal standard.

Toward the end of the nineteenth century, the common law contributory negligence regime was criticised for being unfair to victims. In particular, when both parties had met their respective standards of care, or both were deemed negligent, the victim could not recover any compensation from the defendant. As a result, legislation was introduced to apportion damages in accordance with fault.

In the English speaking world, apportionment was introduced first by the US Congress in several federal safety and employment statutes in the early 1900s. In the 1970s and 1980s, most states in the United States either enacted apportionment statutes or adopted the principle judicially. When California adopted the principle judicially in 1975, the court acknowledged that apportionment added substantially to the burdens on the courts. Since more material facts need to be established in order to apportion damages, the costs of each litigation will be higher under an apportionment regime. The court nevertheless decided that the difficulties could be overcome. At the time of the California decision, 25 states in the United States had enacted statutes to replace contributory with comparative negligence.

In England, the Law Reform (Contributory Negligence) Act 1945 provided for apportionment of damages. In Australia, the apportionment of damages was introduced in a series of statutes in South Australia (1936), Western Australia (1947), Queensland (1952), Tasmania (1954), the Australian Capital Territory (1955), the Northern Territory (1956), Victoria (1958), and New South Wales (1965). In New Zealand, apportionment of damages was introduced in the Contributory Negligence Act 1947, although the position from 1974 was changed again by the Accident Compensation Act 1972.

When the California court was choosing between different negligence regimes, it suggested that the assignment of fractions of damages to different parties was not difficult "in theory". It turns out, however, that it is extremely difficult, if not impossible, to assign fractions of damages in a way that gives parties appropriate incentives to take care. It is difficult enough to determine the implications of different rules of apportionment for incentives to take care.

When there is no way for the system to converge on desirable rules for apportioning damages, the rules are likely to vary greatly from one trial to the next. Tort cases based on apportionment rules may become little more than gambles. Victims may be encouraged to try their luck even when their case appears weak. The result has been encouragement for judges and juries to consider other criteria, such as the financial wealth of the parties, in awarding damages. The resulting 'deep pockets' judgments distort incentives to take care. Those likely to be found negligent will be encouraged to invest too many resources in taking care, while those unlikely to be found negligent will tend to be careless.

The defects of apportionment are exacerbated by the award of punitive damages, or penalties greatly in excess of the harm actually incurred. Punitive damages can play a useful role in a system in which legal cases can establish more efficient new rules, for example, to delineate new property rights or to change the legal standard of care in response to a change in the technology for taking care. The incentive for any one individual to bring legal action may then be far less than the social value of a more efficient delineation of rights or responsibilities. Punitive damages can raise the private benefits of litigating and make them closer to the social benefits. If damage awards are unrelated to expected costs and benefits, however, punitive damages may simply exacerbate the costs of perverse incentives.

Administrative costs are also likely to be much higher under an apportionment regime. Trials will be more expensive under such a regime, and the number of litigable cases will be greatly increased relative to a common law contributory negligence regime. Under apportionment, almost everyone remains at least partially liable for every accident, opening the way for many more law suits. The rule of negligence was introduced in part to limit the number of law suits that otherwise occur when everyone remains liable to everyone else. With apportionment we have, however, come full circle. Now almost everyone is at least partially liable to almost everyone else. High legal costs are probably the only reason an accident is not litigated in an apportionment regime.

When individuals are risk averse, it might be thought that the sharing of damages that occurs under apportionment provides an offsetting benefit. Apportionment, however, is an inefficient method of providing insurance. Danzon (1991: 52) reports that the overhead on \$1.00 of compensation using the tort system has been estimated at 120% as compared with 20% for large group insurance programmes. Similarly, Rogers (1994: 31) reports that the 1973 (Pearson) Royal Commission on Civil Liability and Personal Injury estimated that the cost of operating the tort system in the United Kingdom was 85% of the value of compensation paid through the system, while the corresponding figure for social security (excluding collection costs borne by the employer) was about 11%. Liability rules are most appropriately directed at providing incentives to avoid injuries, leaving insurance to provide compensation to victims.

The introduction of apportionment by legislation reflects several defects of democracy. Apportionment was introduced because it was thought to be fairer than the common law rules of strict liability and negligence. In particular, apportionment was designed to achieve *ex post* outcomes that more closely reflect the relative care taken by the parties involved in an accident. While the common law rule of contributory negligence appears to favour injurers in terms of *ex post* outcomes, it favours victims in an *ex ante*, and less visible, sense by encouraging injurers as a group to exercise too much care. A complete view of the equity implications of a negligence rule can only be obtained by examining its effects on the distribution of expenditures on care, in addition to the distribution of damages in cases that come to trial. Legislators inevitably focus on the immediate and obvious effects at the expense of the longer-term and more indirect consequences. The increase in litigation under an apportionment regime is probably also not an unwelcome development from the perspective of the lawyers. Perhaps it illustrates the tendency for democratic politics to favour vested interests at the expense of the public.

CONCLUDING REMARKS

Efficiency of resource use ought to be a major concern of society, as it is the principal means by which material living standards may be improved. We should help the poor, and such aid is most efficiently provided through private welfare. Welfare financed through taxation should be a measure of last resort. Governmental activities other than welfare should generally focus on achieving greater efficiency, although we have also noted instances where freedom and equality before the law ought to take precedence over efficiency.

Efficient resource use encompasses an appropriate balance between present use and conservation of resources for future use. Arguments in favour of the claim that markets will not encourage the efficient use of resources over time are mistaken. Such arguments ignore the opportunities that markets provide for people to buy and hold resources in the hope of making capital gains as resources become more scarce. The capital accumulation, technological progress and rapid economic growth in economies in which markets play a major role also belie the claim that markets lead to a devaluation of the interests of future generations.

Arguments suggesting that markets undervalue future resource use also fail to compare markets with the relevant alternative. Political resource allocation mechanisms, even in democracies, have not promoted efficient resource allocation in general, nor have they safeguarded the interests of future generations. Politicians often ignore outcomes occurring after the next election. The accumulation of debt over the last half of this century and the deterioration of public infrastructure over recent decades suggest we should not rely heavily on the wisdom and beneficence of government in order to secure our future.

The excessive exploitation of some resources, adduced as evidence that markets devalue the interests of future generations, may in fact be evidence of an inadequate reliance on markets. Resources that are not owned tend to be overexploited and abused in a ‘tragedy of the commons’. Defining and allocating property rights to such resources would allow them to be traded in markets. Owners would have an incentive to trade-off the value of current versus future use.

Attacks on the inheritance of property are inconsistent with the claim that people ought to be encouraged to preserve more resources for future use. The focus of these attacks is a concern for inequality of opportunity. A more appropriate notion of equity of opportunity would, however, focus on people who are severely disadvantaged regardless of the cause. There is no reason to redistribute to people who are not poor. In particular, membership in some identifiable subgroup of the population does not justify making exceptions to the principles of horizontal equity and equality before the law.

Finally, the common law legal system has been corrupted by attempts to use it to redistribute income. Laws encapsulate general rules of behaviour thought to be desirable on average and over the long term. The legal system is not well suited to making exceptions, and is best targeted at promoting the goal of efficiency. Redistribution on the grounds of compassion is more effectively handled by explicit transfers to the poor.

8

CONCLUSION

Equity is a complex notion that incorporates both what we ought to do as a society and what we can do. The drive toward a more equitable social order involves the creation of rules and institutions that will be fair to all. In this chapter we suggest how our arguments have addressed the questions posed in the Introduction: How can we decide what the goals of public policy ought to be? Why should efficiency and liberty be included among our social goals? Ought equality or some other notion of equity be a social goal? How might equity best be incorporated into policy?

HOW CAN WE DECIDE WHAT THE GOALS OF PUBLIC POLICY OUGHT TO BE?

Ethical principles are necessary, but not sufficient, for determining social goals. Analyses of the likely consequences of different actions or institutional arrangements are also needed. Good intentions often lead to undesirable results.

In Chapter 2, we discussed two popular ethical positions, the deontological or Kantian view, and the utilitarian view. On the Kantian view, the morality of a given action does not depend upon the consequences of that action, but merely upon the intention of the actor. Before acting, each person should be able to will that the rule justifying this action become a universal law, applying to all who find themselves in a similar situation.

Utilitarianism is, by contrast, a consequentialist ethical theory. We should, on the utilitarian view, consider the consequences of our actions before we act. We should act so as to maximise the happiness of society. We considered two types of utilitarianism: act and rule utilitarianism. There are problems with each version of the theory.

A pure act-utilitarian could justify violating the rights of one or a few individuals in order to promote the happiness of a greater number of people. Act utilitarianism may also be impossible to implement in practice. The more indirect and delayed effects of one's actions may be overlooked even though they are more important when considering the effects on society as a whole.

A rule-utilitarian maintains that one ought to follow rules that generally promote maximising the happiness of society. At any moment, however, the rule-utilitarian must be able to abandon the proposed rules, if the consequences of doing so will be beneficial. If a rule-utilitarian forgoes the option of abandoning rules, then the rule-utilitarian is not a utilitarian at all, since a utilitarian considers the consequences of an action before acting. A rule-utilitarian who always follows the rules is, in effect, a Kantian.

Neither Kantianism nor utilitarianism accommodate the problems that beset people seeking a complete and effective moral code. In practice, most people rely on ethical rules to guide their behaviour, but they remain open to modifying those rules in the light of experience. Perhaps as a result of such an iterative process, utilitarianism and deontological theories often lead to similar ethical rules.

Kantian-type beliefs that are widely shared may be more persistent than principles based only on utilitarianism. When a person has a conviction that an ethical rule is right, and that conviction is shared by many others in the society, a great deal of evidence will be needed to dissuade that person from acting on the basis of the ethical rule.

Our relationship with others is dealt with in ethical theories by postulating a 'social contract' among people. The social contract is a convenient legal fiction that provides a bridge between the normative and the positive aspects of government. As a matter of fact, people in modern Western democracies adhere to ancient contracts, laws and precedents because it is traditional to do so. A strictly conservative position that opposes all change in principle is, however, untenable in a world in which almost all aspects of society other than its rules and institutions are altering rapidly.

The iterative way in which ethical rules develop is reflected in the social arena. To gain acceptance by a majority of the population, social reforms and institutional innovations need to be more evolutionary than revolutionary in nature. Nevertheless, we must decide how to evaluate new policy proposals or institutional innovations.

As a practical matter, it is useful to suppose that each citizen has agreed to a social contract, to be bound by a set of rules to which all would agree when forming a new society. While no explicit social contract was ever signed, the notion that each of us implicitly agrees to obey certain general social rules is a beneficial convention. Deciding what is appropriate in general, or on average, rather than what happens to be of immediate benefit to oneself, is a hallmark of moral behaviour. Social contract theory encourages people to abstract from their own particular circumstances when speculating about appropriate social rules or institutions.

The Rawlsian conception of a decision-maker behind a "veil of ignorance" is a modern version of social contract theory. A person behind the veil of ignorance is said to be choosing the preferred rules and institutions for the society in which that person will live. Such a person is presumed not to know, in Rawls's words, "his place in society, his class position or social status, nor does any one know his fortune in the distribution of natural assets and abilities, his intelligence, strength, and the like". Empirical evidence is essential, however, to postulating the choices of people behind a veil of ignorance. We need to consider common features of preferences and facts about the available technologies, such as the type of goods or services people like to consume, the likely cost of producing those goods and services, and the degree of risk aversion people generally exhibit.

WHY SHOULD EFFICIENCY AND LIBERTY BE INCLUDED AMONG OUR SOCIAL GOALS?

Kantian principles guide our belief in moral absolutes and our affirmation of the importance of individual liberty, horizontal equity, and the application of a set of rules equally to all. Utilitarianism leads to support for efficiency, but may also encompass what Hume termed a 'sympathy' for the pain of others, an inability to ignore the hardships others face, and a desire to help people in desperate need.

From a utilitarian perspective, the pursuit of efficiency is the most important goal of society. An allocation of resources is said to be efficient if resources cannot be reallocated in a way that would allow the people who gain to be better off after compensating any people who lose to ensure that the latter are no worse off than they were originally. Part of the normative appeal of efficiency as a social goal is that an inefficient allocation of resources represents, in a sense, waste.

For a rational self-interested individual behind Rawls's veil of ignorance, efficiency corresponds to the potential average living standard. The only circumstances under which someone would not opt for a higher average living standard would be when that higher living standard is necessarily accompanied by other characteristics that the decision-maker finds undesirable.

In more practical terms, policies are endorsed or rejected under the efficiency criterion based on whether they can be expected on average to lead to desirable or undesirable consequences for people. In particular, according to the efficiency criterion, it does not matter who gains or loses from a change.

Rational self-interested individuals contemplating a social contract are also likely to be concerned about the distribution of benefits resulting from economic activity. Many economists temper their support for efficiency by recognising that respect for individual liberty and some notion of equity may also be important determinants of the desirability of different policies. Even when other goals are being pursued, however, efficiency remains relevant as a measure of how much prosperity the society is forgoing on average in order to further other goals.

In most countries with democratic political institutions and market-oriented economic ones, and particularly in the English-speaking world, there is also a strong Kantian-type belief in the value of individual liberty. Voluntary cooperation is preferred to coercion. In particular, the coercive power of governments ought to be used only as a last resort to achieve benefits that cannot be achieved voluntarily. People should be free to pursue their own moral or religious values or cultural practices provided they do not infringe on others' rights. By the same token, the coercive power of the state ought not to be used to promote particular moral, religious or cultural values.

From an economic perspective, a concern for liberty may often be compatible with policies or institutions that maximise efficiency. For example, when people are free to make voluntary exchanges with each other they could be expected to trade until the potential gains are fully exploited. More generally, we usually decide whether a particular set of institutions is likely to lead to an efficient outcome by reasoning about how individuals will behave in response to incentives and constraints.

We discussed two situations in which the pursuit of liberty and efficiency may conflict. The first occurs when one person cares about the action of another person and not just the latter's well-being. The former may then prefer the latter to behave differently. For example, a donor may prefer that a recipient spend donated funds on food, shelter, education or medical care while the recipient might prefer to use the funds for entertainment and travel. An efficiency criterion would take the preferences of the donor into account. A liberal would, however, respect the preferences of the recipient.

Government coercion of people to pay taxes is another case in which liberty and efficiency may conflict. Such coercion infringes upon liberty, but may promote efficiency. A society based on markets and individual freedom would be impossible without a monopoly enforcer of the rule of law. Resources are needed, however, to enforce the law and also to provide services such as defence against common external threat. If contributions toward providing such public goods were made voluntary, individuals would have an incentive to free ride on the generosity of others. A certain amount of coercive political action may raise efficiency, even if taxation is an infringement of liberty.

On the other hand, it may be difficult to obtain accurate information about preferences for jointly provided services when people are coerced into paying for them or have them supplied free of charge. Allowing citizens to vote for competing decision-makers in periodic elections, and allowing free speech, may provide alternative avenues for eliciting preferences regarding jointly provided services. As we discussed in Chapter 3, however, such mechanisms are a poor substitute for markets when markets can be expected to achieve a tolerably efficient outcome.

Even when liberty and efficiency are not in conflict, there may be another practical difference between the two goals. The public appears more committed to the principle of liberty than to efficiency, perhaps because liberty is less abstract and is instantiated in freedom of expression and freedom of religion. Thus, efficiency may in practice be promoted most effectively by emphasising the value of liberty as a social goal.

OUGHT EQUALITY OR SOME OTHER NOTION OF EQUITY BE A SOCIAL GOAL?

Some scholars argue that equalising income ought to be an essential part of the social contract. This goal may be placed alongside efficiency, liberty and other goals with a balancing of the various goals when they conflict. For Rawls, greater efficiency is of no value unless it raises the welfare of the person likely to be worst off. Among the class of ethical theories in which both efficiency and equality are valued, Rawls's position places maximum weight on equality. (For a detailed explanation of this claim, see the appendix to Chapter 3.)

The notion that increasing equality in the distribution of money income or wealth is a worthy social goal gained more widespread acceptance in the twentieth century. While one now rarely hears that socialism is desirable as an engine of economic efficiency, policies of extensive government intervention and control of economic processes are still defended on the grounds that they promote greater equality. We have presented a number of arguments in support of our position that equalising incomes is a mistaken goal for policy.

Rawls proposed that a rational person behind the veil of ignorance would pursue a maximin strategy of choosing a set of social institutions that maximised that person's worst conceivable level of welfare. We showed that this claim is tenable only if the individual is presumed to be infinitely risk averse, something that is at odds with what we know about human nature. Even a person who is extremely risk averse, although not infinitely so, would choose to move away from a situation of guaranteed equality, as long as the average standard of living could thereby be increased. We also showed that a very risk-averse person anticipating future inequality is nevertheless likely to take on new risks provided those risks raise average living standards and are not highly correlated with existing risks. Finally, we discussed empirical evidence suggesting that positive skewness in the distribution of living standards (roughly speaking, a greater chance of high incomes than of low ones), as well as higher average levels, can compensate for increased risk. In fact, risk-averse people can seek out risky opportunities that have a negative effect on expected outcomes, if the positive skewness is high enough. A person behind the veil of ignorance with commonly observed attitudes to risk would not opt for a situation of guaranteed equality.

While equality is not a desirable end state, a rational but risk-averse person behind the veil of ignorance may be supportive of redistribution. A reduced standard of living in those situations in which a person ends up being well-off may be more than compensated by a higher standard of living in situations in which a person would be poor. If we associate living standards with income, we might conclude that, because people are risk averse, higher taxes paid when income is greater could be more than compensated for by lower taxes, and possibly income or in-kind transfers, when income is lower.

The tax and welfare system is, however, an extremely expensive way of providing such income insurance. The administrative costs and efficiency losses of taxes and transfers greatly exceed the administrative costs of private insurance. Private insurance contracts include large deductibles when administrative costs are high relative to the amount of insurance provided. The analogue in the case of tax and welfare policy would be catastrophic insurance that provided payments for people who are destitute, but did not redistribute income across a broad range of the income spectrum.

Government income redistribution through taxes and transfers is not the only way of insuring people against fluctuations in income or expenditure. Markets can, at much lower cost, insure against many causes of income or expenditure fluctuations including accidents, fires, natural disasters, diseases, disabilities and death of family members. Personal savings can also be used to smooth fluctuations in income or expenditures.

There is a substantial body of economic literature investigating the lifecycle and permanent income hypotheses. The permanent income hypothesis expresses the idea that consumers base their expenditures more on their normal or expected incomes than on their current incomes. When the consumer is experiencing temporary good luck, most of the extra income is saved. Conversely, when a consumer earns an income that is temporarily below what could normally be expected, consumption is not immediately reduced. The consumer instead draws upon savings, or borrows against future income, in order to maintain consumption.

The lifecycle hypothesis augments the permanent income hypothesis by postulating that people do not expect either their income or their desire for market goods and services to remain constant over their lifetime. Most people expect to earn little when they first enter the work force, with their income increasing as they gain more experience, and finally tapering off as they near retirement. Similarly, most people expect their desire for market goods and services to be greatest when they are paying off a mortgage and have children at home to be fed, clothed, educated, entertained and provided with health care. The reaction of current consumption to income fluctuations would then not be exactly as the permanent income hypothesis predicts but will instead depend on the part of the lifecycle in which the recipient now exists. The two theories give very similar predictions, however, for very short-term income fluctuations.

Empirical investigations of the permanent income hypothesis have found that most families are permanent-income consumers. The major exceptions are families with very low holdings of liquid assets and low current incomes. Families with low holdings of liquid assets do not have the savings to support their former consumption levels, while credit constraints apparently prevent them from borrowing against any future anticipated improvement in their incomes. The spread of credit cards throughout the general society appears to have diminished the role of credit constraints to some extent, but there is also substantial evidence that credit constraints remain significant for many low-income families.

Low current income, low holdings of liquid assets and inadequate access to credit do not necessarily condemn a family to poverty. Studies in the United States have found many such families are able to maintain their expenditures. Evidently, they receive private income transfers from extended family members, friends or private charities that are not recorded in formal income measures.

The extensive private options that people use to spread income fluctuations over time have another implication. Government taxes and income transfers often redistribute from a person or family to that same person or family at some other time. In effect, taxes and transfers are a very expensive, and often unnecessary, alternative to the private options available to most people. The main counter-example involves low-income families who are currently credit constrained. Governments should provide a form of poverty assistance for credit-constrained families, but not the extensive redistribution across many income levels proposed by advocates of greater equality.

Our discussion of the rules likely to be included in the social contract focused on the distribution of living standards. The discussion of income redistribution as a form of insurance focused on the distribution of money incomes. The distinction between a standard of living and income suggests another objection to the pursuit of greater income equality as a goal. Money income, or even consumption of market goods and services, is at best an approximate measure of living standards. People value much more than money and what it can buy. For example, wage rates have to compensate people for living in less attractive environments. Moreover, some people are happier in jobs that pay less, but that allow greater flexibility or more interesting work. People also value friendships and family ties, making it difficult to attract them to a new job at any wage the employer is willing to pay. Since income is not a good measure of living standards, redistributing to reduce income inequality could exacerbate inequality in living standards.

A weak correlation between income and living standards is less of a problem when the objective is to alleviate poverty rather than to reduce inequality. The correlation between income and living standards is likely to be much higher at very low income levels, since few families with a low money income are able to maintain an acceptable standard of living. Furthermore, if poverty is defined at a sufficiently low income level, it may be reasonable to err in providing assistance to people who do not need it in order to ensure that those who do need help are accommodated. There is little justification, however, for attempting to equalise measured annual money incomes across a broad range of income levels.

Regardless of the serious conceptual problems involved, many economists have used annual money incomes to measure inequality. We discussed several measures of income inequality that have been proposed, including the coefficient of variation, the Lorenz curve, the Gini coefficient, and Atkinson's measure. For each of these measures, a situation is deemed to be more equal (and better) if people with higher incomes suffer a reduced income, even if other members of society realise no monetary benefit from that reduction. The only benefit accruing to those who are not wealthy is that they now have less reason to be envious of the rich. In practice, redistribution aimed at reducing inequalities between middle- and higher-income groups is likely to reduce everyone's standard of living, including the standard of living of the poor. Inequalities among people, each of whom has a decent standard of living, are not important enough to justify the large efficiency costs of redistribution.

While we reject the notion that equality of wealth or income ought to be a social goal, horizontal equity (or equal treatment of equally situated people) and equality before the law in particular are worthy goals. A set of legal and political institutions is horizontally equitable if it treats like individuals in the same way. Discriminatory laws or policies that differentiate between people on the basis of irrelevant characteristics, is undesirable. Equality before the law also focuses on processes rather than outcomes. An equitable arrangement of social institutions allows people to pursue their own self-interests, provided they do not infringe on others' rights to pursue their goals. Laws apply equally to all members of society and no person is above the law.

HOW MIGHT EQUITY BEST BE INCORPORATED INTO POLICY?

Discussing whether income equality ought to be a goal of policy in principle is not the same as discussing whether it ought to be a goal of democratically elected governments in practice. It makes no sense to argue that we ought to do something that cannot in practice be done. Thus, one also needs to consider the likely consequences of giving democratically elected governments the right to redistribute income or implement other kinds of policies.

Unfortunately, the electoral process is an imperfect mechanism for achieving joint social goals. The information available to voters is deficient relative to what is needed in order to vote rationally. Public discussion of policy tends to focus on the intentions behind policies rather than the actual consequences, on resource inputs rather than outcomes achieved, on immediate and obvious consequences rather than longer-term or more indirect effects.

Democratic governments often pander to vested private interests at the expense of a wider majority interest. A voter has weak incentives to become informed, or to participate actively in influencing governmental decisions, unless the policy concerned is of immediate relevance to the voter's own welfare.

In order to succeed in advancing private interests at the expense of the public, the consequences of special interest policies cannot be too conspicuous. The policy must appear to serve some wider interest that is acceptable to a majority of voters. Unfortunately, there is no guarantee that these wider interests have any moral worth. While some might believe that the majority is always right, we are all aware of past evils associated with such a conviction. Politicians can appeal to widely held beliefs such as racism, xenophobia, sectarianism or religious fanaticism just as easily as they appeal to appropriate ethical values.

Many government policies ostensibly motivated by compassion for the poor, such as public education, often appear to serve the interests of middle-income voters, who often are the marginal (or swing) voters. Voters are often unaware, however, of the amount of additional taxes they have to pay to support programmes such as public education or public health care, or of the efficiency losses caused by those taxes.

It is also difficult to appreciate the efficiency losses associated with a lack of competition in providing services such as education. Private suppliers find it difficult to compete with a publicly supplied service financed through taxes, even when the publicly supplied product is of low quality. Customers of the public supplier do not pay a fee that depends on use. Hence, if they switch to a private supplier, they bear the full cost of using the new service while receiving only a trivial share of any reduction in taxes permitted by reduced expenditure on public education. Furthermore, while services such as public education are often supported on the grounds that they promote equality, in truth poorer families often must use the low quality, publicly provided service.

We argued that democracy involves a tension between serving the desires of a majority and serving private vested interests. Vested interests are often the main beneficiaries of policies such as supplying education through a monopoly public agency. A major reason that services such as public education deliver low value for money is that they cost more than they should. Suppliers use the monopoly status of the public provider to extract rents in the form of wages, staffing levels, or working conditions that could not persist in a more competitive environment.

If we want to help the poor by providing them with services such as public education or health care, we should subsidise provision from competitive private suppliers. The market for services such as education and health care could be made more competitive, and the outcome more efficient, if middle- and higher-income families were left to purchase these services in exchange for paying lower taxes. Subsidies in the form of vouchers or direct payments to suppliers ought to be restricted to the poor.

Poor people might be better served by receiving cash rather than being required to consume particular services in order to receive assistance. Subsidising the consumption of services can at best benefit the recipients by as much as a money transfer of equal cost. One argument against giving cash to the poor is that the parents may spend the money on themselves and neglect their children. Thus, assistance might best be a combination of income transfers and vouchers to subsidise the consumption of particular services such as vocational training, education or child health care.

Many public spending policies that are supported because of their ostensible benefit to the poor actually provide greater benefits to people who are not poor. Arguments used to support the progressive tax system suffer from a similar credibility gap. Taxes that are progressive in the marginal rate are often supported on the ground that they are more equitable than a flat tax which imposes the same marginal rate on everyone. We showed in Chapter 6, however, that progressive marginal tax rates are more equitable than constant marginal tax rates only if it is desirable to redistribute from taxpayers who currently have a high income to taxpayers currently experiencing a middle income. If concerns about equity are motivated by compassion for those who are currently poor, as opposed to envy of those currently experiencing a high income, a more desirable policy would be to couple a flat tax, with a single marginal rate, with income transfers to the poor.

We also argued in Chapter 5 that policies based on envy are unlikely to succeed in their objective and may even be counterproductive when measured against their ostensible goal. If we remove one basis for envy, another is likely to replace it. Conversely, if the number of dimensions along which people can compare themselves with others can be limited, the feelings of envy might actually become more rather than less acute. When people can compare themselves on many dimensions, status becomes an ambiguous concept. Different rankings of people are possible depending on which dimension is regarded as of paramount importance. With only one dimension of comparison, however, rankings are less ambiguous and envy may be more widespread. Redistributing to assuage feelings of envy reduces prosperity and compromises individual liberty for a cosmetic reason.

Government income support as a measure of last resort for those in desperate circumstances is not cosmetic. Such support reflects a compassionate concern for people in need, not envy of those who are better off than we are. We have a duty to help people who are suffering. While it is best to rely primarily on families, friends, and private charity to provide such assistance, government assistance for those who have failed to benefit from other sources of support is also desirable. Redistribution targeted at assisting the poor would be achievable at a lower cost in terms of reduced efficiency than the extensive welfare state we have now.

People who are below or near a defined poverty level ought to be the focus of government income transfers. The United States Panel on Poverty and Family Assistance suggested that the poverty level "should represent a budget for food, clothing, shelter (including utilities), and a small additional amount to allow for other needs". The expenditure bundle constituting needs is a matter for debate and public discussion. Periodic adjustment of the bundle through an explicit public process would encourage the necessary debate. Such a procedure is far preferable to an automatic indexing of a previously defined set of expenditures to average weekly earnings or even to the consumer price index, which includes many goods or services in the consumption bundle that could not be regarded as necessary expenditures to avoid poverty. Periodic debates about the appropriate definition of poverty would also have the beneficial effect of emphasising that the appropriate goal of policy is poverty assistance and not reducing inequality.

In summary, equity, or fairness, can best be achieved by safeguarding each individual's right to personal safety and personal property, by ensuring that each individual is equal before the law, by promulgating a belief in the power and duty of families and private charities to help the indigent, and by creating a limited government welfare programme to aid those in need who fail to receive familial or charitable assistance.

BALANCING SOCIAL GOALS

In our personal behaviour, we attempt to balance adherence to absolute moral rules against utilitarian considerations. We prefer situations in which there is no conflict between what our principles suggest and what is expedient. When a conflict between these goals exists, we follow absolute moral rules when we believe this will serve our best interest in the long run. One reason we follow rules is that our own behaviour affects the behaviour of others in ways that can be either beneficial or harmful to us. If we break a rule in order to secure an immediate benefit, others are likely to follow our example.

There are a number of desirable social goals, such as the pursuit of liberty, equal treatment of equals and equality before the law, that are similar to moral principles. The pursuit of efficiency, and the desire to assist people living in poverty, have a more utilitarian foundation. Social decision-making involves a continual balancing of various goals rather than the exclusive pursuit of any one of them. While a representative democracy is the only practical way of allowing the average person to contribute to this decision-making process, it is far from perfect.

Constraints on the actions of a majority in a democracy, such as limiting redistribution to measures based on compassion alone, can be viewed as analogous to the moral constraints we place on our personal behaviour. Such constraints might limit the good that the majority can achieve in the short run. By altering the way people interact with each other and with governmental authority, however, these constraints might allow society to achieve more in the long run.

Index