Queen's Economics Department Working Paper No. 983

REDISTRIBUTION

Robin Boadway
Queen's University

Michael Keen
International Monetary Fund, University of Essex, UK and the Institute for Fiscal Studies, UK

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

3-1999

REDISTRIBUTION

February, 1999

Robin Boadway* and Michael Keen†

Abstract: This paper, prepared for the *Handbook of Income Distribution* (edited by A.B. Atkinson and F. Bourguignon), reviews some of the central issues that arise in thinking about the motives for, politics of, constraints on and measurement of, redistribution. Amongst the themes are: the potential usefulness of apparently inefficient policy instruments in overcoming the self-selection constraints on redistribution and limiting the damage that ill-intentioned policymakers can do; the continued (perhaps increased) ignorance as to the effective incidence of many key taxes and benefits; and, while there are circumstances in which redistribution may plausibly generate efficiency gains, the likelihood that some trade-off between equity and efficiency is inescapable.

JEL Classification: D3 D6 D7 H2 I3

* Queens University, Kingston, Ontario, Canada K7L 3N6; *E-mail:* boadwayr@qed.econ.queensu.ca

† International Monetary Fund, Washington D.C. 20431, United States; University of Essex, UK; and the Institute for Fiscal Studies, UK. *E-mail:* mkeen@imf.org

Acknowledgments: Helpful comments and advice from François Bourguignon, Sanjit Dhami and Jeremy Edwards are acknowledged with thanks. Views expressed are not necessarily those of any institution with which we are affiliated.

1 Introduction

One good reason for being interested in the distribution of income, the topic of this *Handbook*, is because one might want to change it. The term itself has become unfashionable in many parts of the world, but much of what governments do is manifestly redistribution of one kind or another. Private individuals also undertake large amounts of redistribution themselves. Whatever the labels they go by, distributional issues are, and will remain, the lifeblood of social and political debate, and of much economic analysis.

This paper reviews the rationale, nature and extent of redistribution, by which we shall simply mean an unrequited transfer of resources from one person to another. This is a vast canvas, so vast that—even in such a long paper as this—many issues are inevitably left unaddressed. Redistribution among regions within a federation is not considered, for example, and nor is redistribution among countries. Nor do we deal at any length with intra-generational redistribution within the family.[1] Some issues we cover only briefly, especially if good treatments are already available. Moreover, the perspective is mostly theoretical: the theoretical advances over the last twenty years or so have been more profound, arguably—and as yet—than the empirical ones.

The paper starts by considering the motives for redistribution (Section 2). Amongst these is the exercise of greed and compulsion, leading to a focus in Section 3 on the politics of redistribution. Section 4 then considers the limits to the redistribution that can be achieved in practice, and Section 5 considers the issues that arise in trying to measure the distributional impact of public sector policies. The concluding section highlights some of the central themes that emerge from this review.

2 Motives for Redistribution

One can distinguish three broad rationales for redistribution:

- the pursuit of social justice as an ethical imperative;
- the achievement of mutually advantageous efficiency gains; and
- the exercise of self-interest through the coercive power of the state.

The conceptual distinction between these motives is less sharp, however, than it may at first seem. The notion that ethically acceptable decisions are those that would

[1] Thus we use the terms 'individual' and 'household' synonymously. The classic reference on the allocation of resources in the family is Becker (1981). For a recent survey, see Bourguignon and Chiappori (1992).

be taken, *ex ante*, behind a veil of ignorance as to one's own position in society, for example, leads readily to the adoption of an objective—maximization of the expected utility of a representative citizen—that is much the same as that to which a naive concern with efficiency would lead. And even if it involves redistributing away from themselves, for those with power to impose the implications of their own ethical preferences on others with different views is no less an exercise of coercive power than is expropriation of resources directly to those in power. In practice, of course, any measure of redistribution will rarely reflect the operation of just one of these motives.

Though the distinctions between them thus quickly become blurred, these three motives nevertheless provide a useful organizing framework within which to assess some of the main approaches to the analysis of redistribution. The rest of this section focusses on the first two broad class of motive, leaving the third until the next section.

2.1 Redistribution as Social Justice

One might view a concern with distributional issues as a matter of ethical preference that transcends—or at least, is not explicitly modeled as part of—the determinants of economic behaviour. It is as if people have two distinct personalities, with their self-interested selves essentially disjoint from their ethical selves, a perspective adopted by Arrow (1951a). Such an approach might be defended simply on the lines of analytical convenience: our self-interested preferences guide our day-to-day participation in the market economy, while our ethical ones apply to our participation in collective decision-making. Such a split personality is made very sharply by the device—introduced by Vickrey (1945), subsequently and independently developed by Harsanyi (1955) and used too by Rawls (1971)—of attaching ethical significance to the choices people would make about social arrangements behind a hypothetical 'veil of ignorance' as to which of all possible people they will become. Even to conceive of such an exercise requires that one be able to dissociate one's self-interested self from one's selfless self, with the latter best equipped to make ethical judgements.[2]

Within this class of approaches, different views are taken as to what constitutes social justice. The most common—still, indeed, the dominant approach to the normative economic analysis of redistribution —is the *welfarist* one (a term due to Sen (1970a)), according to which the social desirability of alternative social states is to be assessed in terms of some Bergson-Samuelson Social Welfare Function $W(U_1, ..., U_N)$ defined over the utility levels U_i, $i = 1, ..., N$ of the population under study. This function is naturally taken to be both anonymous (permuting utilities across individuals has no effect on social welfare) and Paretian (non-decreasing in all arguments). Tighter

[2]Some would argue that ethical behaviour (honesty, altruism, etc.) can be viewed as being ingrained through an evolutionary process akin to the biological evolution of physical traits. For an outline of this approach, see Frank (1989).

ethical judgements, and restrictions on the kind of interpersonal utility comparisons one is willing to make, impose further structure on the form of one's $W(\cdot)$: see, for example, Boadway and Bruce (1984).

The potentially powerful implications of the welfarist approach were recognized by Edgeworth (1881) over a century ago, and later stated most generally by Dasgupta, Sen and Starrett (1973). Suppose that all N members of the population have the same strictly concave utility function $U(Y)$ defined over lump sum income Y. Then the maximisation of any S-concave social welfare function—meaning, roughly, one that is increased by replacing each individual's utility by a weighted average of all utilities[3]—requires setting $Y_i = (1/N)\sum_k Y_k$, $\forall i$: that is, complete equalisation of incomes.[4] Note that the case for redistribution of *income* that thus emerges is quite distinct from any concern with the distribution of *well-being*: one arrives at complete equalisation of incomes even in the utilitarian case $W = \Sigma U_i$, in which government cares not about the distribution of utility but only its total. What ultimately drives redistribution in this case is not so much a concern for equality of utilities as a diminishing private marginal utility of income combined with identical utility functions.

This welfarist approach lends itself well to the applied analysis of distributional issues. Each individual—and, more particularly, each policymaker— can be conceived of as having a distinct social welfare function, and perhaps too distinct cardinalisations of utility. For any given preferences, and given a complete specification of how the economy functions, it is then in principle a straightforward matter to compute an optimal distribution of income.

For marginal reforms, in particular, welfarism provides an elegant and practicable analytical framework. Differentiating with respect to the lump sum income of the ith person gives a distributional weight $\beta_i \equiv (\partial W/\partial U_i)\partial U_i/\partial Y_i$ that enables one to compare and aggregate real income gains to different units: the higher is β_i the greater is the social gain from an increase in i's real income, and redistributing \$1 from individual k to individual i raises social welfare if and only if $\beta_i > \beta_k$. These distributional weight have proved a convenient way of introducing distributional concerns into applied welfare analyses. These have been widely used in cost-benefit analysis (Drèze and Stern (1987)) and have enabled an appealing approach to the identification of desirable tax reforms. Ahmad and Stern (1984), in particular show how the restriction that $\beta_i > 0$ for all households i can be used to identify Pareto-improving reforms; Mayshar and Yitzhaki (1995) show how the further restriction that the β_i fall with income enable reforms to be identified that increase social welfare for all concave $W(\cdot)$. Results of this kind hold the prospect of narrowing the range for dispute over sensi-

[3]More precisely (and introducing the convention of indicating vectors by bold type) an S-concave function is one such that $F(P\boldsymbol{x}) \geq F(\boldsymbol{x})$ for all bistochastic matrices P (all square matrices, that is, with $P_{ij} \geq 0, \sum_i P_{ij} = \sum_j P_{ij} = 1$, $\forall i, j$.

[4]This of course ignores the possible incentive effects of taxation; Section 4 looks at the limits to redistribution these create.

ble directions of reform, since all can subscribe to the same judgement even though they may differ in the precise form of their social preferences. Methodologically, this approach of exploring the implications of social welfare functions that appear in the analysis as something of a *deus ex machina* sits well with economists' preferred view of themselves as giving dispassionate advice on emotive issues: it enables one to analyse the consequences of alternative distributional judgements without endorsing any.

Though it has thus proved empirically valuable, welfarism has powerful critics. One line of attack sees it as too permissive in attaching no significance to the process by which the economy reaches its final state. No importance is attached, for example, to any property rights that might stand in the way of redistribution. An extreme libertarian position would regard as illegitimate—a denial of social justice rather than a potential contribution to it—redistribution that violates property rights that have been legitimately acquired. Such a view, associated with Nozick (1974), leaves little scope for redistribution. Nor, conversely, is any independent importance attached to equality of opportunity, a dimension explored by Roemer (1996) and Fleurbaey and Maniquet (1998). Even in a minimalist state, however, distributional issues cannot be entirely avoided: for the basic functions of the nightwatchman—such as maintaining law and order—create efficiency gains that must somehow be shared across the members of the economy (Inman (1985)).

The second line of criticism of the standard welfarist approach, urged most forcefully by Sen (1985, 1992), sees it as too narrow: it focusses only on individuals' utilities, and moreover sees these as reflecting only real incomes in a somewhat narrow sense. It is possible, for example, to imagine redistribution going hand in hand with political repression (as perhaps in some formerly communist countries); one might wish to attach importance to the latter as well as to the former. A range of social concerns beyond material well-being—Rawlsian 'primary goods'— may also matter. One such is liberty; and Sen (1970b) shows very starkly how even the apparently minimally liberalist notion that individuals should be allowed to make the final decision on some issues, such as what books they themselves read, may lead to Pareto inefficient outcomes.[5] Importance might also be attached, in and of itself, to whether people deal honestly with one another, or to whether they behave corruptly. Moreover, what matters in evaluating alternative economic arrangements may not be income in the usual sense but some rather wider notion of the quality of life enjoyed. This underlies Sen's notion of capabilities and functionings. The idea is that poverty is not simply an inadequacy of income; it is a lack of basic capabilities to function as a human being. Functioning involves being able to take part in activities such as clothing, feeding and housing

[5]Sen's celebrated example concerns two persons, prudish person 1 and racier person 2, and their joint decision as to whether 1,2 or neither should read *Lady Chatterly's Lover*, called options x, y and z, respectively. Person 2's preference ordering is $z > x > y$, while 1's is $x > y > z$. Thus, x Pareto dominates y. But, Sen argues that liberalism would suggest that z is preferred to x (1 should not be forced to read the book), while y should be preferred to z (2 should be allowed to read it); so, y should be preferred to x on liberal grounds, which contradicts Pareto domination.

oneself, participating in community activities, appearing in public without shame, avoiding morbidity, and so on. Alleviating poverty then involves generating minimally acceptable capabilities, which in turn requires that resources be tied to personal circumstances (such as age, gender, health status) and social circumstances (physical and social environments, epidemiological factors, public health characteristics). This approach has been developed into an alternative to the standard approach to distributional analysis described above. The relative difficulty of measuring capabilities, however, has meant that this alternative approach has not yet had the impact on practical distributional analysis that the standard distributional weights have done.

These two—the welfarist and capabilities approaches—are of course not the only possible approaches to social justice and its distributional consequences, though they are currently the most influential. One alternative with a strong historical pedigree, for example, is the notion that all should suffer equal sacrifice in financing government, a precept discussed fully in Musgrave (1959). Though this does not in itself call for redistribution—no sacrifice is needed if there is no public expenditure to be financed—it clearly is a distinct principle of distributive justice. And it has some strong implications. With such a view of the world it is far from obvious, for example—even in the absence of incentive effects—that taxes ought to be progressive.[6]

2.2 Efficient Redistribution

One of the ideas most deeply ingrained in thinking about economic policy is the notion of an inherent trade-off between equity and efficiency. Efficiency-enhancing policies increase the size of the economic pie; redistribution, on the other hand, cuts the pie more fairly but in the process causes it to shrink.[7] This stark contrast view is reflected in Musgrave's (1959) influential distinction between three roles of government: stabilization, allocation and distribution. Allocative efficiency and distributive justice were seen as inherently conflicting objectives, and for clarity of purpose, at least, best kept distinct.

In recent years, however, there has been increasing interest in the possibility that redistribution and efficiency might not always conflict, but that in some circumstances, it might be possible to have both: to make the pie larger by cutting it more fairly. At a policy level, this new view of redistribution was lent special force by the observation—

[6]It is straightforward to show, for example, that equal absolute sacrifice in financing some fixed level of public expenditure (requiring $U(Y - T(Y)) - U(Y) = \kappa$ for all Y and some κ) implies an increasing average tax rate if and only if the income elasticity if the marginal utility of income exceeds unity in absolute value. (This follows on setting the derivative of the utility difference to zero, multiplying by Y and rearranging to find that $T^{'}(Y) > T(Y)/Y$ iff $U^{'}(Y)Y$ is decreasing in Y).

[7]Having succumbed to one cliché, we should for completeness refer to another: Okun's (1975) notion that money transferred from rich to poor is carried in a 'leaky bucket': "[s]ome of it will simply disappear in transit." (p.91)

at least until recently—of economic success going hand in hand with improved distributional equity in some of the tiger economies (Kanbur, this volume, chapter 13). At a theoretical level, it was fed too by a recognition that in a second-best world redistributing income might act on additional constraints in such a way as to generate efficiency gains. This section explores the possibility that redistribution might actually improve the efficiency with which the economy operates.

There is a semantic point of some importance here. The term 'efficiency gain' is used in the literature in at least three different senses, to mean: a Pareto improvement; a potential Pareto improvement, meaning that all could gain if appropriate compensation were paid; an increase in total output, or some other such measure of aggregate activity. An efficiency gain in the first sense clearly implies also a gain in the second, but beyond that the links between the three notions of efficiency gain are tenuous. Clearly too each has rather different ethical status. All, however, are potentially of interest.

2.2.1 Altruism

One case in which redistribution can obviously be Pareto-improving is that in which transferors are altruistically inclined towards recipients (early recognition of this is in Sen (1967), Hochman and Rogers (1969) and Thurow (1971)). Indeed such altruism carries the possibility that some redistribution — though not, perhaps, enough — will be undertaken voluntarily. Historically, charity has been a major source of support of the poor, and voluntary giving remains sizeable in many developed countries: around 2 to 3 per cent of GNP in the US, for example. Altruistic donations are perhaps even more important in developing countries, being a major source of support within families, extended families and even communities. Much of the international aid flowing from developed to developing countries presumably reflects altruism, at least in large part. More important for our purposes, altruism has potentially profound implications for the need for, and the effectiveness, of redistribution by the public sector.

The simplest view of altruistic redistribution sees the well-being of the recipients as a public good from which the donors derive pleasure. Imagine, for example, an economy comprising a single poor person, whose welfare $U^p(C_p)$ depends only on their consumption C_p, and N rich individuals, each of whose welfare $u^r[C_r, U^p(C_p)] \equiv U^r(C_r, C_p)$ also depends positively on that of the poor. To see the scope and limits of voluntary redistribution in such a context, consider the properties of a symmetric Nash equilibrium in which each rich person i chooses their gift to the poor, g_i, so as to maximise their own utility $U^r(Y_r - g_i, Y_p + g_i + G_{-i})$, taking as given the contributions $G_{-i} \equiv \sum_{k \neq i} g_k$ of all others; denote the solution by $g_i^* \geq 0 \ \forall i$.

This non-cooperative equilibrium is indeed Pareto-superior to that without redistribution: the poor person is clearly better off, and so too is each rich person (they have

revealed themselves to benefit by contributing, and benefit also from the gifts of the other rich).

But the equilibrium is also easily seen to be Pareto inefficient: all, rich and poor alike, could gain if each rich person were to give slightly more. This inefficiency is an immediate implication of viewing the income of the poor as a public good, for there then arises a standard free-rider problem: each rich person ignores the benefit that their gift to the poor conveys to other rich people, and so each ultimately gives too little.

This inefficiency creates scope for a Pareto-improvement through compulsory redistribution.[8] There is, however, an important limitation on what can be achieved: in this model, a limited amount of redistribution is liable to prove entirely ineffective, having no effect on the final well-being of the poor. The reason is straightforward: each rich person cares only about their own consumption and that of the poor person, and their ability to offset a redistributive tax levied on them by reducing their own voluntary transfers means that such a tax has no effect on the opportunity set they perceive. To see this formally, denote by T the tax imposed on each rich person, proceeds of which are redistributed to the poor. Imagine that all rich people except the ith respond to the introduction of this tax by reducing their gifts by the full amount of the tax, to $g^{**} = g^* - T$. The problem of the ith rich person is then to choose g_i to maximise $U^r(Y_r - g_i - T, Y_p + g_i + \sum_{k \neq i} G^{**}_{-i} + NT) = U^r(Y_r - \tilde{g}_i, Y_p + \tilde{g}_i + \sum_{k \neq i} G^*_{-i})$, where $\tilde{g} = g_i + T$. This is of exactly the same form as the problem in the absence of intervention, and so has the solution $\tilde{g}_i = g^*$, implying $g_i = g^*_i + T$: thus if all others reduce their gifts by the amount of the tax, so too will person i.

This ineffectiveness of redistribution clearly requires the altruistic link to be operative for all the rich: the inability of the rich to extract transfers from the poor means that compulsory redistribution which is taken so far that the voluntary gifts of the rich fall to zero will indeed affect the final income of the poor. The implication is powerful, nevertheless. Pareto improving redistribution does not merely require public intervention: it requires enough redistribution to crowd out private transfers. A little redistribution may have no effect.

This result bears an obvious resemblance to Barro's (1974) version of Ricardian equivalence in the sense that it predicts the neutrality of government transfers. But there is a major difference: in Barro's model, the allocation dictated by the dynastic utility function of the 'representative' agent is fully efficient. (See Piketty (this volume, Chapter 8)) for an explicit treatment of the dynastic model.) Altruism is kept within the dynastic family, so that the effects of all transfers are fully internalised. There is then no reason to regret the ineffectiveness of public policy: there is no inefficiency

[8]The discussion focuses on the direct transfer of resources; an alternative response to the inefficiency is to subsidize private gifts, as often indeed often occurs through the tax deductibility of charitable donations.

to be addressed. Inefficiency re-emerges, however, with a recognition that the socio-biological facts of life render the Barro dynastic family untenable: each child has parents from two separate families, so that all families are inextricably linked through future inter-marriage. The implication, developed by Bernheim and Bagwell (1988), is that a parent caring for his or her own offspring indirectly cares for the descendants of all contemporary parents. Thus, the altruistic bequest again becomes like a public good, and its level of provision inefficient. Too little saving will be done for future generations (perhaps none) a point that was recognized some time ago by Marglin (1963) and Sen (1967). Though this in turn suggests a case for public intervention, policy is again liable to be ineffective. Pursuing the argument to its limits, Bernheim and Bagwell (1988) provide a general statement of the considerable power of altruism to generate policy ineffectiveness. In particular, the ineffectiveness of policy continues to apply —until gifts cease to be made—if transfers to the poor are financed not by lump sum taxes on the rich but by distorting taxes on their labour supply (see their Example 2); and it applies too if the government makes no transfers to the poor but instead subsidises the gifts of the rich (Andreoni (1990)). The ineffectiveness of redistribution in models of altruism as enjoyment of a public good is in this sense rather general. To have effect, public redistribution must drive out private charity.

Returning to the model of redistribution as contribution to a public good, two implications for the situation of the poor deserve emphasis.[9]

The first is that the poor may find that any increase in their own pre-tax income Y_p is largely offset by a reduction in the gifts that the rich are willing to make to them. For they face an implicit marginal tax rate whose magnitude is determined by the preferences of the rich. In the limit, if the income elasticity of demand of the rich for the net income of the poor is zero,[10] then the poor face an effective marginal tax rate of 100 per cent. In effect, the rich then care only about raising the poor to some minimum level of income, and then tax away any excess that the poor acquire by reducing their gifts. Such a situation gives rise to the *samaritan's dilemma* (Buchanan (1975)): knowing that their donations might induce the poor not to earn income, should the rich nonetheless promise to bring the poor up to some minimum level? We return to this question below.

A second striking implication is that an increase in the number of the rich can actually leave the poor worse off: not only may the net income of the poor deviate further from its first-best level (from the perspective of the rich) as the number of the rich increases—reflecting the intensification of the free-rider problem—the gift of each rich

[9]Both follow from routine comparative statics on the necessary condition for g_i, evaluated at a symmetric equilibrium,

$$U_Y(Y-g,Ng)+U_G(Y-g,Ng)=0\,,$$

together with the second order condition $U_{GG}-U_{YY}<0$. It is assumed too that the income elasticities of demand for both private consumption and consumption of the poor are strictly positive.

[10]So that $U_{YY}-(U_Y/U_G)U_{GY}=0$.

person may actually fall, and moreover by so much as to more than offset the increase in the number of people making the gift.[11]

The view of altruistic redistribution as the enjoyment of a public good thus has extremely strong implications. Some, indeed, appear to be inconsistent with practical experience. The model also predicts, for example—the well-known theorem of Warr (1983), further developed by Bergstrom, Blume and Varian (1986)—that a redistribution of income amongst rich donors will have no effect on the final income of the poor; this runs counter to empirical evidence cited by, for instance, in Hochman and Rogers (1973). The model also implies that all rich contributors will enjoy the same level of utility, irrespective of any difference in their private incomes.[12]

So apparently implausible are the implications of this simple model of altruism that there has been much interest in developing alternatives. One such is to suppose that donors derive utility not only from the well-being of the recipient but also from the act of giving itself. In this view the ith rich person has utility $U^r(C_r, C_p, g_i)$, the appearance of g_i reflecting the *warm glow* of giving. This formulation blunts many of the stark implications of the public good approach: public redistribution will not be fully offset even when gifts are operative, for example, because the rich are no longer indifferent between \$1 that they give to the poor themselves and \$1 that the government gives; and the implicit marginal tax rate on the poor will be lower, since even if their gift becomes less needed the rich will be reluctant to forgo the distinct pleasure they derive from giving. Further implications are developed by Andreoni (1990).

The picture of redistribution as motivated by a selfless altruism is in many ways a happy one. So beguiling is it, indeed, that its lack of intrinsic ethical worth deserves emphasis. For example, in Section 4.5 we briefly review the literature on redistribution in federal systems, due originally to Boskin (1973) and Pauly (1973). This grounds redistribution in the supposition that the rich in each local jurisdiction are altruistically inclined towards the mobile poor; in one contribution that we shall later focus on, Wildasin (1991), for instance, characterizes policies that are 'optimal' in the sense of maximising the welfare of the rich. But it is all too easy to forget—as Wildasin does not—that this is a narrow and conceivably repugnant sense, there being no reason for wanting to order society according only to the wishes of the rich, however well-intentioned they might happen to be. Indeed in a wider context altruism can clearly have a dark side. It may be racially or sexually biased. The altruism of the rich and

[11] This is the case, for instance, if $U_{YG} = 0$ and initially $N = 1$ (here treating the number of rich people as a continuous variable).

[12] Intuitively, since all enjoy the same consumption of the poor, and the income elasticity of demand for the net income of the poor is strictly positive, the rich can all be content with their gift only if they all have the same net income. (More formally, the result can be seen by differentiating the necessary condition (A.1) with respect to Y to find that the equality can hold for distinct net incomes only if $U_{YY} - (U_Y/U_G)U_{GY} = 0$.) See Itaya, de Meza and Myles (1997).

powerful towards their own children may perpetuate inequalities of opportunity. The task of policy then quite possibly becomes not to facilitate the exercise of altruism but, on the contrary, to prevent it.

Note too that what looks like altruism may reflect a sharp self-interest: transfers may be made to the poor to forestall violence or theft (Sala-i-Martin (1997)).

2.2.2 Social Insurance

Some people have bad luck. That bad luck can take many forms, including becoming unemployed or disabled, having poor health, having more children to finance than expected, having fewer children to look after one than planned, choosing a feckless partner; the list is depressingly long. If they have not been able to arrange appropriate insurance themselves, such ill-fortune creates an evident case for redistribution—from the lucky to the unlucky—on the grounds of social justice encountered earlier. Moreover, it may be that the prospect of such *ex post* redistribution will raise the expected utility of all, including those who in the event will prove to be lucky. In this sense social insurance may prove efficiency-enhancing. Indeed, as was noted long ago by Vickrey (1945) and hinted at above, one very general kind of bad luck—being born with low intrinsic ability to earn—can be seen as providing a very general basis for systematic redistribution across differing income levels: behind the veil of ignorance, people would prefer a system of redistribution amongst the living that would effectively insure them against the risk of being born with low ability.

Insurance market failure

The question, of course, is why individuals are unable to purchase such insurance for themselves. This may seem evident enough in the case of innate ability, since the outcome of the uncertain event is presumably publicly revealed at the time an individual reaches sufficient maturity to enter contracts. Even in this case, however, parents may be able to purchase insurance on their offsprings' behalf. For the narrower risks, prominent amongst the reasons for potential failure of insurance markets are asymmetries of information between insurance companies and those wishing to buy insurance. These asymmetries are conventionally grouped into two kinds: adverse selection (those wishing to buy insurance being better informed as to the likelihood of the event occurring to them than the insurance company can be) and moral hazard (insurance companies being unable to observe actions taken by the insured that affect the likelihood of the insured peril occurring or the size of the damage). The precise implications of these asymmetries are typically quite model-specific, sensitive to such matters as the conjectures insurers have as to whether actions they take which make contracts offered by others unprofitable will result in those contracts being withdrawn. In the broadest terms, however, the implication is that insurance markets may fail to exist, and if they do exist may reach equilibria that could be improved upon by a

government with no better information than is available to private companies.[13]

Consider, for example, a possible market in which individuals can buy insurance against the possibility of becoming unemployed. Suppose that there are two types of workers, differing only in the probability of losing their job. Figure 1 illustrates. In the absence of any insurance, each consumes at point α, enjoying a wage of w when employed and non-labour income of κ if not. The two types have the same (von Neumann-Morgenstern) preferences, and are risk-averse, so that their level curves of expected utility are convex to the origin as drawn, and differ only in so far as the probability of unemployment differs between them: specifically, the indifference curves of the group with the higher risk of being unemployed are steeper (the intuition being that since they are more likely to find themselves unemployed they are willing to forgo more consumption when employed for a marginal increase in consumption when unemployed). The line $\alpha\alpha$, the fair-bet line for the low-risk persons, shows the set of points at which the expected consumption of a low risk individual would be the same as at the endowment point α. Thus it represents the set of contracts which would exactly break even if taken up only by the low risks, with all points below that line — offering less consumption to the policy-holder in both states of nature — making strictly positive profit if taken up only by the low risk. Similarly, $\alpha\beta$ is the set of contracts that break even if taken up only by the high risks, while $\alpha\gamma$ is the set that would break even if taken up by all.

FIGURE 1

In an ideal world, each type of consumer would be able to fully insure themselves against unemployment. If an individual's type were costlessly observable, one would indeed expect competition to lead to this outcome, with contracts offered at both A and B: low risks buy the former, high risk the latter, and neither faces any consumption uncertainty. Suppose, however, that risk-type now becomes unobservable. Then all high risks will try to buy contract A; but that means both types are buying A, and since A lies above $\alpha\gamma$ the contract makes a loss. Adverse selection thus prevents competitive markets attaining a first-best outcome.

What then will be the laissez faire outcome when insurance companies cannot observe type? The answer depends on the assumed form of market conduct. Suppose, following Rothschild and Stiglitz (1976), that each firm takes as given the set of contracts offered by other firms and can observe the total quantity of insurance each individual purchases. It is then easy to see that there cannot be a 'pooling' equilibrium in which both types buy the same contract. If there were, for profits to be zero, it would have to be at a point like C in Figure 1(a). But since the indifference curve of the low risk

[13] The material in this subsection is largely standard, and dealt with at more length by, for example, Atkinson (1991), Barr (1993) Hirshleifer and Riley (1992) and Mas-Colell, Whinston and Green (1995).

type through this point (IC'_L) is flatter than that of the high risk (IC'_H) there exists an area like that shaded in the picture such that any contract in this area would be purchased by the low risk but not by the high and, being below $\alpha\alpha$, would make a profit.

The only possibility is then of a separating equilibrium in which the two types buy different contracts, illustrated in Figure 1(b). Such an equilibrium must constrain the contract for the low risks to be no more attractive to the high risks than the contract designed for the latter. The high risks are fully insured at point B. The contract bought by the low risks must then be along αE so that it is not bought by the high risks. But if the zero profit line for pooled contracts is $\alpha\gamma$ there can then be no equilibrium: any contract in the shaded area shown would attract the low risk away from E (or anything to the left of it) and the high risks away from B; and, being below $\alpha\gamma$, any such contract will make a profit. In this case there thus exists no equilibrium. In other cases, however, there does exist a separating equilibrium. Suppose, for example, that the zero profit line for pooled contracts were $\alpha\gamma'$. Then the shaded area vanishes, and contracts B and E do indeed constitute a separating equilibrium.

What then are the implications for public policy? While governments, by virtue of being able to elicit information from households in a mandatory way, may become better informed than private insurers, the interesting case is that in which government is no better informed. When it exists, the above separating equilibrium is constrained efficient: a general result in models of this sort. But if no equilibrium exists, a Pareto gain can be obtained by introducing a compulsory scheme of pooled insurance with all persons required to purchase a common amount. Redistributive objectives will have been satisfied since the unlucky will benefit at the expense of the lucky. But, this will be partly at the expense of a system in which low risks are cross-subsidising high risks.

On the other hand, if, unlike in the above model, private insurers are unable to monitor individual purchases of insurance (say, because they can spread purchases among more than one insurer), a pooling equilibrium can exist. Higher risk persons will purchase more insurance than lower risk ones at the common premium. In such a setting, compulsory provision of insurance in equal amounts to all can be implemented so that the lower risk households are better off, while high risks can be better or worse off (Johnson (1977)).[14]

It is thus conceivable that a redistributive scheme of social insurance can generate an efficiency improvement. The same may be true if the state is able to reap scale

[14]This result depends critically on the ability of households to choose their quantities of insurance freely. Dahlby (1981) shows that if a Rothschild-Stiglitz equilibrium exists, a necessary, but not sufficient condition for compulsory insurance to be Pareto-improving is that households be free to purchase voluntary supplements. Moreover, if an equilibrium in the sense of Wilson (1980) exists, compulsory insurance can never be Pareto improving.

economies in the provision of insurance (a point which, as Atkinson (1991) reminds us, was given some importance in the Beveridge report), if the costs of administering a compulsory state scheme are less than those of a competitive scheme (as argued by Arrow (1963)) or if there are economies of scope in running insurance schemes of this sort in tandem with the tax collection system.

Much of what governments do—pensions, unemployment and sickness benefit—does indeed look so like the provision of compulsory insurance, with benefits payable at similar rates to all and subject to a test not of current means but rather of past contributions, that it is naturally labeled social insurance. Such measures are now often rationalised by reference to the kind of informational arguments above. Quite how persuasive an interpretation this is, however, remains unclear. Atkinson (1991) notes, in particular, that such models typically have the feature that it is the low risks who fare worst under market provision, just as in the example above it is they that bear some residual risk, with the high risks being completely insured. This seems a puzzle, since casual observation suggests that the concern which has often driven compulsory insurance schemes has been that it is the high risk that are hardly done by, not the low risk. This suggests that more work is still to be done on the relationship between insurance market failure and state provision. Perhaps—though this is not the case in the model above—it is precisely the adverse externality experienced by the low risks in the laissez faire equilibrium that galvanises them into supporting compulsory state provision. Or perhaps it is the case, for example, that the high risks also face a liquidity constraint that prevents them taking out the full insurance that would otherwise be available to them, so confounding the potential inefficiencies of insurance markets with capital market imperfections of the kind discussed below.

Risk-taking and investment

The model just analysed was one of pure adverse selection: people had no economic decisions to make. In practice, they face a myriad of choices that affect both their future incomes and the risks to which they are exposed. And a key concern in discussion of the welfare state has been precisely with the likely impact that the provision of insurance will have on such decisions. Will it lead people to save too little, for instance, and will it induce them too take too much risk (knowing there is a safety net to catch them if they fall)?

The potential trade-off between the desire to provide insurance and the desire to preserve appropriate incentives to work and save was raised and explored by Eaton and Rosen (1980) and Varian (1980). Sinn (1995, 1996) emphasises that redistribution may have beneficial effects on risk-taking. To see the considerations at work, it is useful to combine both sets of concerns into a single model.[15]

[15]This is essentially an extension of that in Varian to capture Sinn's concerns by allowing for a decision on the extent of risk-taking.

Consider then a two-period world in which a large number of (*ex ante*) identical individuals must take two decisions before the uncertainty is resolved. For one, they invest an amount x in acquiring human capital, yielding them $x+\lambda\theta$ in period 2, where θ is stochastic with (for simplicity) all realisations negative.[16] Clearly the choice of x will be affected by, but does not in itself affect, the uncertainty faced. The second choice variable does: by spending an amount e on risk prevention activities in the first period—studying harder or longer to enter a safer profession (a doctor rather than an actor) perhaps—the stochastic component of second period income is affected through $\lambda(e)$, with $\lambda^{'} < 0, \lambda^{''} > 0$. Note that e affects both the uncertainty of second period income, its variance being $\lambda^2 \mathrm{var}(\theta)$, and expected second period income, $x + \lambda(e)\bar{\theta}$ (where $\bar{\theta} \equiv E[\theta]$). Preferences are described by a von Neumann-Morgenstern utility function $u(C_1) + v(C_2)$ over consumption in the two periods, with both u and v increasing and strictly concave. For definiteness, it is further assumed that $v''' = 0$.

If individuals could insure fully, all would optimally share equally in second period income. Assume a large number of identical individuals (their number normalised to one), then aggregate second period income is $x + \lambda\bar{\theta}$, and the first-best problem reduces to that of maximising $u(w - x - e) + v(x + \lambda(e)\bar{\theta})$, with necessary conditions

$$x : -u^{'} + v^{'} = 0 \tag{1}$$

$$e : -u^{'} + \lambda^{'}\bar{\theta}v^{'} = 0 \tag{2}$$

To explore the effects and optimal design of redistributive taxation as a response to imperfections of the insurance market, suppose now, at the opposite extreme, that individuals are for some reason unable to buy any insurance. The only insurance is that implicitly provided by the government, which taxes second period income at rate t and returns the proceeds $t(x+\lambda(e)\bar{\theta})$ as a poll subsidy; there are no taxes or subsidies in period 1. People take the poll subsidy as independent of their actions, so that the typical individual's necessary conditions are:

$$x : -u^{'} + (1-t)E[v^{'}] = 0 \tag{3}$$

$$e : -u^{'} + (1-t)\lambda^{'}E[v^{'}\theta] = 0, \tag{4}$$

which implicitly define investment and risk-avoidance decisions as functions $x(t)$ and $e(t)$ of the tax rate.

The no-intervention equilibrium—that with $t = 0$—is clearly inefficient. Specifically, it can be shown that the level of risk-prevention e is higher in the absence of intervention than it is in the first-best:[17] the absence of insurance makes people too cautious. It also

[16] It may seem odd that the return to human capital is negative. But this is only for simplicity: it would be straightforward, but notationally burdensome, to ensure a strictly positive return in all states.

[17] To see this, note first that (1) and (2) give

$$\lambda'(e^F)\bar{\theta} = 1 \qquad (F.1)$$

leads them to invest too little in their human capital: x is lower in the no-intervention equilibrium than in the first-best.[18]

Consider then the effects of the redistributive tax scheme. Solving (3) and (4) for $x(t)$ and $e(t)$ and substituting into lifetime expected utility gives the effect of a tax increase on maximised expected lifetime welfare $V(t)$ as:

$$\frac{V^{'}(t)}{E[v^{'}]} = -\lambda(e)\frac{\mathrm{cov}[v^{'},\theta]}{E[v^{'}]} + t\bar{\theta}\lambda' e' + tx'. \tag{6}$$

This shows clearly the three effects at work. The first is an insurance effect: the presence of the tax reduces the uncertainty of net income, since the government cushions some of the loss if things turn out badly but also shares in the gain if they turn out well. Since $\mathrm{cov}(v',\theta) < 0$, this effect of increasing the tax rate is unambiguously beneficial. Indeed if this were the only effect at work it would be optimal to set the tax rate at 100 percent. But there are also the induced effects on x and e to be considered. Note that these matter only for their effects on revenue: since individuals make their investment and risk-avoidance decisions to maximise their own welfare, the only welfare effect of policy-induced changes in these variables is through the external impact on redistributed tax revenues.

Thus the second term in (6) captures the effect of redistribution-cum-insurance on risk-taking. Given a further assumption that $u'' = 0$, it can be shown[19] that $e'(t) > 0$. This effect is therefore beneficial whenever $t > 0$. The intuition follows from the earlier observation that the absence of insurance tends to lead to an excessively high e. Social

---

the superscript indicating the first-best. Noting that (3) and (4) imply $E[v'] = \lambda' E[v'\theta]$ when $t = 0$, using the definition of a covariance and rearranging one finds

$$\lambda'(e(0))\bar{\theta} = 1 - \lambda'(e(0))\left(\frac{\mathrm{cov}(v',\theta)}{E[v']}\right) < 1 \qquad (F.2)$$

From (F.1) and (F.2), $\lambda'(e(0)) > \lambda'(e^F)$,and it follows from the convexity of $\lambda(\cdot)$ that $e(0) > e^F$.

[18]Evaluating (3) at $t = 0$ gives

$$\begin{aligned} 0 &= -u'(w - x(0) - e(0)) + E[v'(x(0) + \lambda(e(0))\theta)] \\ &= -u'(w - x(0) - e(0)) + v'(x(0) + \lambda(e(0))\bar{\theta}) \qquad (F.3) \\ &< -u'(w - x(0) - e^F) + v'(x(0) + \lambda(e^F)\bar{\theta}) \qquad (F.4) \end{aligned} \tag{5}$$

the second equality using the assumption that $v''' = 0$, and the third the observation that the right of (F.3) is decreasing in e combined the result above that $e(0) > e^F$. Comparing (F.4) with (1) gives

$$-u'(w - x^F - e^F) + v'(x^F + \lambda(e^F)\bar{\theta}) < -u'(w - x(0) - e^F) + v'(x(0) + \lambda(e^F)\bar{\theta})$$

and the conclusion that $x^F > x(0)$ follows on noting that $-u'(w - x - e) + v'(x + \lambda\bar{\theta})$ is decreasing in x.

[19]By somewhat involved comparative statics on (3) and (4), making use of the assumption that $v''' = 0$.

insurance can mitigate a tendency towards excessive caution that insurance market failure might otherwise create. The third effect in (6) is that on the acquisition of human capital. This is beneficial, again when $t > 0$ if social insurance tends to counteract the tendency towards under-investment in the absence of insurance. The sign of $x'(t)$, however, is uncertain.

Even in this simple model, with all its restrictions, there are few simple results on the nature of the optimal social insurance scheme: one cannot even show that the optimal tax rate is strictly positive. Balancing the insurance effects of social insurance schemes against their impact on risk-taking and incentives to work and save is a daunting task, as indeed experience makes all too clear.

One implication of the impact of social insurance on risk-taking is of particular interest. Insofar as it encourages risk-taking, an increase in the tax rate may actually increase post-tax inequality: the variance of net income, $(1-t)^2\lambda(e(t))$, may be increased more by the induced reduction in e (and consequent increase in λ) than it is reduced by the direct impact of a higher t. Indeed in Sinn's rather different model this must be true at an optimum so long as that optimum has the feature that more inequality raises average income (an eminently plausible feature, since otherwise by reducing inequality one could increase average income, suggesting that one is unlikely to have started at an optimum). Social insurance, commonly thought of as a response to underlying inequality, may thus serve, perfectly reasonably, to increase it.

The samaritan's dilemma and time consistency

Alternative explanations for some types of social insurance rely on time consistency arguments. One such applies the samaritan's dilemma argument to the public sector. Suppose that society has as a policy objective coming to the aid of the least well-off members of society. How well-off one is in the future depends in some measure upon one's actions undertaken at the present, such as saving to finance future consumption and unexpected contingencies, taking preventative measures to mitigate possible adverse outcomes, and human capital accumulation. If society cannot commit not to provide support to those who encounter bad luck, then people will undertake too little in the way of self-insurance to protect their future well-being. In these circumstances, public provision of insurance can be welfare-improving.

An early formal treatment of the issues is that of Bruce and Waldman (1991). They construct a simple model with a representative poor person with no initial income, and a representative rich person. In the first period, the rich person transfers part of their income to the poor through the government to satisfy an altruistic urge. The poor can consume part and invest part, where the investment can be interpreted generally to include human capital investment or financial investment. Second-period consumption by the poor then depends upon the investment undertaken in the first period as well as further transfers provided by the rich. If the government acting for the rich can commit to a sequence of transfers announced at the beginning of the first period, the poor will

choose an efficient amount of investment in the first-period and redistribution itself will be Pareto-optimal. But, if such commitment is not possible, the poor will have an incentive to under-invest in the first period in the expectation that in the second-period the government cannot avoid making an altruistic transfer to make up for the adverse consequences of low first-period investment. The time-consistent (sub-game perfect) equilibrium with cash transfers will be inefficient: either the poor will simply underinvest in the first period and obtain an excessive transfer in the second period, or the rich will ensure that second-period altruism is non-operative by transferring and consuming an excessive amount in the first-period. But, Bruce and Waldman then show that if the government is able to make part of its transfers in the form of investment in the first period, full efficiency can be obtained even in the absence of commitment.

Coate (1995) extends the Bruce-Waldman argument by allowing private altruistic transfers to exist alongside government ones. In this case, even if the government can fully commit to second period policies, the fact that private donors cannot might still induce the poor to under-invest or, as in the Coate analysis, under-insure for the future. (The argument relies on private altruism being operative enough, despite there being a potential free-rider problem if there are a large number of rich). Again, in-kind transfers of income insurance can restore efficiency.

Similar arguments can be used to justify mandatory public pensions (Kotlikoff (1987)) and mandatory education and training (Boadway et al (1996)). It may be, indeed, that it is the samaritan's dilemma which provides the most persuasive rationale for some of the major transfer schemes observed in developed countries (Stiglitz (1998)).

In the case of unemployment insurance, the time inconsistency of government policy can take another form. If the probability of unemployment is given, private insurers might be able to satisfactorily insure against the chances of individual workers being laid off (excepting for the purposes at hand the possible difficulties of coping with moral hazard problems). But, the probability of unemployment is not randomly determined; it is at least partly under the control of government. In these circumstances, governments might find it irresistible, given private unemployment insurance, to ratchet up the unemployment rate thereby exploiting private insurance companies to the advantage of workers. (This is analogous to the central bank being unable to commit to keep inflation rates low.) Given this, private unemployment insurance markets may not exist, and governments may have to become the suppliers of unemployment insurance (Boadway and Marceau (1994)).

2.2.3 Dynamic inefficiency and intergenerational redistribution

We have already encountered one sort of efficiency argument for intergenerational transfers, that arising from altruism for the well-being of future generations. As with

static altruism arguments, intergenerational altruism gives rise to a free-rider problem, though one which is somewhat more subtle that the static case. For, even if one cares directly only for one's own heirs, indirectly one cares for the heirs of all of one's contemporaries because of the chain of connectedness that arises through the intermarriage of descendants in the future. As we have seen, while this can render small amounts of public intergenerational redistribution ineffective, efficiency improvements can come about once public transfers are large enough to crowd out private bequests.

There is, however, a potential further source of inefficiency in a dynamic setting that might be addressed by redistributive transfers, and that is dynamic inefficiency. The argument is as follows. Suppose income per capita grows at the rate g per period. An on-going intergenerational transfer from younger to older cohorts which yields an implicit rate of return of g to each cohort on the taxes paid when they are young can be sustained in the long run. If g exceeds the market rate of return, say, r, instituting such a transfer makes all cohorts better off as long as the plan stays in effect, and as long as g continues to exceed r. Should such a situation exist in perpetuity, such a transfer will be Pareto-improving. Thus, an economy such that $g > r$ forever is said to be *dynamically inefficient* (Starrett (1972)). Conversely, economies with $r > g$ are dynamically efficient: any intergenerational transfer scheme will make some cohorts better off and some worse off.

The possibility of using intergenerational transfers to achieve dynamic efficiency gains has naturally led to considerable interest in the literature. But, the policy consequences are likely somewhat limited. First, evidence suggests that for much of recent history, rates of return on investment r have exceeded rates of growth in per capita incomes g (Abel, et al (1989)). Second, even if g did exceed r, it would have to do so forever for dynamic inefficiency to be established.[20] Obviously, that is something that no one can predict. Third, the dynamic inefficiency result applies only in an infinite-horizon economy. If, for any reason, one views the economy as having a finite horizon, the phenomenon of dynamic inefficiency disappears: for the last generation is required to make a transfer to its predecessors but has no successors to receive a transfer from itself.

Once the prospect of dynamic inefficiency is discounted, intergenerational transfers become like movements along an intergenerational Utility Possibilities Frontier. Some persons are made better off, and some worse off, and more conventional rationales for redistribution apply: social insurance/justice motives come to the fore. In a static setting, social insurance was designed to consummate a form of insurance against bad luck at birth that households were not in a position to purchase for themselves. In an intergenerational context, it is recognized that not all bad luck at birth is 'insurable' across a single cohort. Some cohorts are unlucky in the aggregate: they may face wars, natural catastrophes, demographic shocks, etc. But, good and bad risks might

[20] Actually, g only has to exceed r on average forever, roughly speaking (Balasko and Shell (1980)).

be insurable over time, that is, across generations. If so, governments could use intergenerational transfers as a policy device for smoothing out real incomes across generations (Gordon and Varian (1988)).

There is some evidence that intergenerational transfers are used to some extent to do so. Wars are financed by at least partly by debt; some major public pension schemes were introduced in the wake of the Great Depression, and so on. However, engineering social insurance schemes across generations encounters obstacles not faced by their static counterparts, which themselves are formidable. Governments last for relatively short periods of time, and there are obvious problems with current governments committing future ones to abide by intergenerational social insurance schemes. At best, the argument for social insurance based on intergenerational risk sharing can provide moral guidance rather than operational advice.

2.2.4 Non-convexities: The case of efficiency wages

We saw in subsection 2.1 above that concavity of the utility function implies that aggregate utility can be increased by equalising lump sum incomes. In the same way, non-convexity of the technology can imply the existence of efficiency gains from redistribution: as first noted by Guesnerie (1975), with increasing returns some competitive equilibria may Pareto dominate others.

FIGURE 2

One especially powerful example of the potential significance of non-convexities is provided by the analysis of land reform in the presence of efficiency wages by Dasgupta and Ray (1986, 1987). Suppose that a worker's effective labour supply is related to their consumption—which for simplicity we take to be identical with their income—by the relation $\lambda(C)$ shown in Figure 2: greater consumption improves productivity, and at low levels does so at an increasing marginal rate. Workers derive this income from two sources: selling their labour at wage rate w, and receiving (lump sum) receipts from their landholdings of κ. If land holdings differ across workers, so too will the efficiency wage: the wage, that is, which minimizes the per unit cost of effective labour supply, $w/\lambda(w+\kappa)$. (Thus the efficiency wage of a landless labourer, for example, is given by the inverse of the slope of the line OA in Figure 2). Since the efficiency wage decreases with the level of land-holding, those with low landholdings are disadvantaged in the labour market. Indeed it may well be that some of those with lowest land holdings are unemployed: their marginal product when paid their efficiency wage may be less than that efficiency wage. For those with large land holding, in contrast, the efficiency wage may be below their reservation wage, in which case the employment decision hinges on the comparison between their marginal product and that reservation wage.

In such circumstances redistribution may lead to an increase in total output. Consider the effects of transferring land from a worker with a large initial land holding, for whom the reservation wage exceeds the efficiency wage, to a worker with a low initial land holding, who is currently unemployed. For the former, so long as the efficiency wage remains below the marginal product there is no change in employment. The induced reduction in the efficiency wage of the poorer worker, however, may be sufficient to bring them into employment. It is thus perfectly possible that redistribution of this kind will increase total output. In practice, egalitarian land reforms do indeed seem, at least in some prominent cases, to have led to an expansion of aggregate output.[21]

Note, in passing, that these results illustrate and emphasize the importance of distinguishing between the alternative notions of efficiency mentioned earlier. For while redistribution here generates an increase in total output it can never generate a Pareto improvement: in attempting to share out the additional output, so as to ensure that all share in the gain, one inescapably undoes the productivity gain that generated it.

The efficiency wage model has one further striking implication for the potential effects of redistributing to the poor, shown by Ravallion (1984). Consider an individual currently paid exactly their efficiency wage. We have seen that redistributing towards such an individual (modelled as an increase in κ) would reduce the wage w that they are paid in equilibrium. It turns out, however, that the wage will actually fall by *more* than the amount of the transfer, so that their total income $w+\kappa$ actually falls.[22] Diagrammatically, the result can be seen by constructing a picture similar to Figure 2, but with w on the horizontal axis. Increasing κ then shifts the λ curve uniformly to the left by the amount κ, and it can be seen that the new tangency with a ray from the origin will imply a wage that falls by more than the transfer increases. Intuitively, the employer could choose to reduce the wage so as to exactly offset the increased transfer; but this reduced wage means that the labour cost saved by an increase in the worker's efficiency is also reduced, so that it benefits the employer to cut the wage still further. Paradoxically, redistributing towards the poor in this case ultimately leaves them worse off.

In the argument above, the non-convexity acted to favor redistribution towards the poor: in the presence of increasing returns, such transfers may help the poor escape from low-level equilibria (a feature which also operates in some of the models of pre-existing distortions that we discuss next). In some contexts, however, non-convexities may point towards redistribution away from the poorest. In the framework above,

[21] Alesina and Rodrik (1994), for example, cite evidence that growth is faster the more equal the distribution of land, and refer to a widespread view that faster growth in Southeast Asia than in South America reflected there having been land reform in the former but not the latter. See also Kanbur (this volume, chapter 13).

[22] Minimising $w/\lambda(w+\kappa)$ requires that $\lambda - w\lambda' = 0$, the second order condition being $\lambda'' < 0$. Hence $w'(\kappa) + 1 = \lambda'/w\lambda'' < 0$.

for example, two poor workers may both have such high efficiency wages as to be unemployed, for example, but a disequalizing transfer of the land owned by one to the other might reduce the efficiency wage of the recipient by enough to make her employable. A similar harsh reality arises in the context of famine relief: minimizing mortality may require a policy of triage, with available resources concentrated on the better off, who stand some chance of survival, rather than shared with those so malnourished as to be beyond hope (Ravallion (1987)).

2.2.5 Capital markets, incentive constraints and other pre-existing distortions

These cases of efficiency-improving redistribution are instances of a more general observation: when there are constraints which prevent the economy attaining a first-best outcome, the utility possibilities frontier may be upward-sloping over some of its range.[23] As a simple example, it is straightforward to show that in a two-person economy with pre-existing commodity taxes it is possible for an unrequited transfer from one to the other to leave both better-off. (Piketty provides an extended treatment of these issues in Chapter 8 below.)

More powerful examples arise in the context of capital markets, which are afflicted by many of the same informational problems as insurance markets. Lenders' informationally-constrained responses to the possibilities that borrowers will default on their loans or simply abscond with the funds, are likely to lead to levels of investment that depart from the first-best. Differing models give differing predictions as to whether there will be under- or over-investment (de Meza and Myles (1987)): lenders may ration investments below the first best, for example, or borrowers may borrow excessively to take advantage of being able to default if things go badly. In either event, however, a redistribution of wealth towards borrowers is likely to move investment closer to the first best, so increasing aggregate income: incentive problems will be mitigated insofar as greater wealth enables (or obliges) investors to rely more heavily on their own funds rather than borrowed funds tainted with lenders' mistrust. Effects of broadly this kind are analysed by Aghion and Bolton (1997) and Piketty (1997). Analogous issues in the accumulation of human capital are addressed by, in particular, Hoff and Lyon (1995)—who shows that lump sum redistribution may be a superior policy to the apparently better-targeted one of taxing or subsidising loans directly—and Galor

[23] Consider for instance the problem of maximising the welfare of person A in a two-person economy, subject only to the constraint that person B achieve utility of at least $\bar{U}$ and the resource constraint $F(\mathbf{x}_A + \mathbf{x}_B)$. By the envelope theorem, the sign of the slope of the utility possibilities frontier is given by that of the multiplier λ in the Lagrangean $L = U_A(\mathbf{x}_A) + \lambda\{U_B(\mathbf{x}_B) - \bar{U}\} + \mu F(\mathbf{x}_A + \mathbf{x}_B)$. From the first order conditions—assuming there is some good i that has strictly positive marginal utility to both individuals—one finds $\lambda = (\partial U_A/\partial x_{Ai})/(\partial U_B/\partial x_{Bi}) > 0$. But adding a further constraint $G(\mathbf{x}_A, \mathbf{x}_B) \geq 0$ means—unless $G(\cdot)$ depends only on the sum of the two consumption vectors—that positivity of λ is no longer assured.

and Zeira (1993). Banerjee and Newman (1993) consider similar effects in relation to occupational choice. Access to one's own resources may also ease the hold-up problem on relationship-specific investments: the reluctance, that is, of partners to finance investments which will leave them, *ex post*, at the mercy of the other party (Hoff and Lyon (1995)). There is an evident link, of course, between capital market imperfections and long run growth, and indeed these have been a focus of much of the recent work in this area. Indeed a recurrent lesson from endogenous growth models has been the potentially beneficial effects of redistribution in improving growth prospects by overcoming capital market imperfections.

The literature now abounds with other cases in which a pre-existing distortion creates a potential for Pareto-gain from redistribution: between two countries in a three-country world (Bhagwati, Brecher and Hatta (1983))[24]; between regions of a federation (Boadway and Flatters (1982)); between countries differing in their productivities in the provision of some international public good (Ihori (1996)).

So pervasive are such examples that it is worth emphasising that even when redistribution leads to an efficiency gain over some range, some trade-off is likely to remain at end of day. Altruism, for example, means that the utility possibilities frontier is upward-sloping over part of its range; but there will generally also be a range of options available on the downward-sloping part. In that sense, it is important to remember, conflict equity and efficiency often ultimately inescapable.

2.3 Redistribution as Expropriation

A third motive for redistribution is simple greed: self-interested individuals may possess, and certainly have an incentive to acquire, the power to redistribute towards themselves. More generally, of course, all practical measures of redistribution, however motivated, emerge from some political arena; and it is to the broader topic of how politics can shape the outcome that we now turn.

3 The Politics of Redistribution

Unless it is Pareto improving, redistribution is ultimately an exercise of coercion and sovereign power and in that sense is an inherently political matter. Indeed Stigler (1970) asserts the understanding of redistribution to be the principal task of political economy.

In primitive societies redistribution is above all a matter of transfers extracted by a ruling elite, with the losers prevented from withdrawing by physical threat. The

[24]The distortion here being the absence of an optimal tariff against some third country.

question is what limits the extent of redistribution in such a kleptocracy. The key point here is that even a self-serving despot can find it optimal to adopt policies that are far from outright appropriation. One reason for this is the threat of insurrection if redistribution towards the elite is taken too far. Thus Usher and Engineer (1987) and Grossman and Noh (1994) characterise despotic redistribution in terms of an extraction of surplus tempered by the possibility of revolution. Another brake on expropriation, emphasised by McGuire and Olsen (1996), is provided by the despot's interest in allowing enough resources to be channeled into productive uses to allow the base of her tax revenues to be sustained. Thus it was that the bandits in the *Seven Samurai* habitually left the villagers enough to sow next year's crop; and thus it was, perhaps, that the slave-masters of the south took a rational interest in the maintenance of their property (Fogel and Engerman (1974)). Indeed, one can view the historical emergence of social insurance as an attempt by entrenched elites to prevent insurgence by the poor.

In what follows, however, we focus on redistribution in societies that have attributes of democracy. Confiscation by physical force is excluded: compulsion is exercised through the rule of law, and those rules have democratic features. We consider in turn three forms of democratic decision-making: democracy in its purest form; representative democracy, bringing in a role for politicians; and the influence of interest groups.

3.1 Direct Democracy

By this we simply mean a situation in which no agent—no politician, bureaucrat, or lobbyist—comes in any substantive way between the preferences of the citizenry and the selection and implementation of policy. Collective decision-making is simply a matter of applying some rule to the expressed wishes of the citizenry. Any politicians or civil servants required to write the policy into law or implement it are merely the faithful ciphers of the citizenry, able to commit to pursuing the policy decided upon and making no more than a reasonable call on resources in doing so.

The questions immediately arises as to the rule that society will, and should, choose for moving from individual preferences to the selection of a collective policy. This is the subject matter of social choice theory, which we make no attempt to survey here: see, for example, Mueller (1989) and, for an account with a view to issues of public economics, Inman (1985). There are many rules that one could imagine being used to make decisions on distributional matters. Unanimity has occupied a central place in libertarian treatment of optimal voting rules. Or one might conceive of a super-majority—a majority, that is, of some number over 50 per cent—being needed for compulsory redistribution. The literature on distributional politics, however, has focused almost exclusively on just one choice rule: decision on the basis of a simple majority. This methodological feature has had such a marked impact on the literature

as to require some discussion in itself.

3.1.1 Majority rule

In their treatise on collective decision-making, Buchanan and Tullock (1962) refer to the selection of a voting rule as the constitutional stage to be evoked, at least conceptually, from behind the veil of ignorance. Citizens will trade off the expected benefits from increasing the size of majority required to pass legislation—reducing the chance that they will be in the losing group when votes are taken on issues with distributive consequences—against the increased costs of decision-making. They envisaged there being some sort of consensus about the size of majority that would trade off these benefits and costs, but what that would be was left unexplained. The case for majority rule based on the Buchanan-Tullock calculus was provided in axiomatic form by May (1952) and later Rae (1969), and recounted by Inman (1985) and Mueller (1989). More pragmatically, the apparent prevalence of majority rule in democracies—albeit as a matter of voting in legislatures rather than of literal direct democracy—provides an obvious reason for interest in majority rule. More loosely, one might invoke majority rule—as do Alesina and Rodrik (1994)—as a simple way of capturing the idea that even despots cannot ignore the mass of opinion. Analytical convenience too, doubtless has some part in explaining the popularity of the approach.

Notice, however, that the mere specification of majority rule does not completely characterize the voting rule, for the question then remains as to precisely who it is that may vote. There may be important technical restrictions on the franchise: most particularly, future generations cannot themselves vote today on policy decisions that must be made today and which can powerfully affect their welfare. Even putting such issues aside, the extent of the franchise is itself a matter for choice and, hence, for explanation. Historically, major extensions of the franchise have been associated with real fear amongst the previously enfranchised that democracy will result in their expropriation: thus De Tocqueville warned of the 'tyranny of the majority,' and Lord Derby saw the 1867 Reform Bill that he and Disraeli brought into being—a substantial extension of the franchise—as 'a leap in the dark.'[25] For present purposes, however, we simply follow the literature in assuming that all economically relevant agents are enfranchised.

Majority rule is thus widely assumed. Modelling it requires some notion of equilibrium. The most natural is that of a *Condorcet winner*: an option which cannot be defeated in a majority vote by any other feasible alternative. The well-known paradox of voting, however, is that there may exist no such option: preferences (and the relative numbers who hold them) may be such that a majority prefer some option a to another b, b to c and yet a not beating, but rather being beaten by, c. Everything being beaten by

[25] Quoted in Blake (1966), p.474.

something else in pairwise votes, the ordering implied by majority rule is intransitive. Under majority rule there is no guarantee that a best element will exist.[26]

There are various ways out of this dilemma (Sen, 1977), such as relaxing the axioms underlying the approach, or allowing the collective decision rule to use richer information than just household orderings. For the purpose of legitimizing majority voting, restrictions on the domain of preferences are most relevant. The literature has sought restrictions on preferences that will preclude such intransitivities. Here there arises an important distinction as to whether one or several dimension of policy issue are at stake.

In the one-dimensional case—meaning that the policy options from which selection is to be made can be described in terms of a single parameter—a well-known sufficient condition[27] for the existence of a Condorcet winner is that every individual's preferences be *single-peaked*: that is, that there be some ordering of options such that each individual has only one local optimum. The Condorcet winner is then easily characterised: ordering voters by the location of their preferred point, the Condorcet winner is the option most preferred by the median voter (Black (1948)).

Establishing that preferences are (or are not) single-peaked is thus the typical first-step of much analysis from a political economy perspective. It turns out that single-peakedness can be problematic even in rather simple models, and examples of this will be encountered below. But the failure of single-peakedness is not necessarily a disaster, for single-peakededness is a special case of a more general sufficient condition for the existence of a Condorcet winner: Sen's (1966) condition of value-restricted, which indicates the extent to which voter preferences must be sufficiently similar to avoid intransitivity. Sen's condition requires that among all combinations of three alternatives, voters must agree that some alternative is not worst, or some is not best, or some is not in the middle.[28] In practice, this may not be an easy property to check. A less powerful condition, but one whose economic content may be more relevant for redistribution issues is the Gans and Smart (1996) notion of *single-crossing* (which, as discussed below, differs from, but is related to, the well-known Mirrlees-Spence single-crossing property). A sufficient (but not necessary) condition for value restriction—and hence a sufficient condition for the existence of a Condorcet winner—is that there exist some ordering of individuals (or, more precisely, preference types) and some ordering of options, such that if one person prefers a to b (respectively, b to

[26]This is but one instance of Arrow's Impossibility Theorem: no collective decision rule based on individual preference orderings alone and respecting certain widely supported properties (including the Pareto principle) will yield a complete and transitive ordering of alternatives for all possible preferences of voters.

[27]Note no preference restriction can in itself be *necessary*: if a majority of the citizenry have the same preferences then majority rule will be transitive however weird the preferences of the minority.

[28]Pattaniak and Sen (1969) formalised this notion of value restriction into the following axiom on preferences: if for any three alternatives x, y, z some individual's preference ordering is $xPyPz$, then any other individual for whom zPx must also have $zPyPx$.

a) then so do all those higher (lower) in the ordering. It is immediate, moreover, that single-crossing in this sense again implies that the median voter is again decisive.

Matters are much less straightforward in the multi-dimensional case. As Plott (1967) showed, there may exist no Condorcet winner even if all have single-peaked preferences along every dimension of policy: see the summary in Atkinson and Stiglitz (1980). There do exist restrictions that will yield transitivity of majority rule—Bucovetsky (1991) e.g.—but they are tight. One would certainly not expect them to be satisfied in contexts of redistribution. The reason is easily illustrated. Consider the problem of sharing a pie of size unity between three individuals A,B and C. The set of possible allocations of the pie is then described by the three-dimensional unit simplex, shown in Figure 3. At point A, for example, individual A receives all of the pie—her bliss point—whilst along the axis BC she receives none. Assuming that individuals care only about their own share, indifference curves correspond to circles around each individual's bliss point. It is readily shown that for any allocation another may be found which defeats it in a majority vote, so voting cycles are endemic. For example, α beats β in a majority vote, and β beats γ; but γ beats α. Cycling is likely to be pervasive in direct voting over redistribution.

FIGURE 3

Reflecting these profound methodological differences, models in which the underlying issue space is multi-dimensional commonly have decisions taken through representative democracy, with citizens voting over candidates rather than policy bundles. With only two candidates, the paradox of voting disappears—at least from immediate view. For the disappearance may be illusory, since there will typically exist no Nash equilibrium in the platforms offered by the two candidates. Intransitivity may thus manifest itself.

Unidimensionality is thus a great convenience in addressing voting problems, though even then single-peakedness is far from assured. For the US, there is some reassuring evidence that politics is a largely uni-dimensional business, particularly in recent years: senators voting records suggest that they line up on issues in roughly the same order (Alesina and Rodrik (1994)). In other historical contexts, however, there is evidence of two or more dimensions being relevant: see, for instance, the discussion in Roemer (1998). In any case, it is clear that, reducing complex distributional issues to one-dimensional problems is largely a matter of brute force.[29]

One dimensional models of this kind have thus been hugely influential in analysing the politics of redistribution; we now review some of the principle applications of this kind.

[29]Alternatively, some of the problems of multi-dimensional voting can be avoided by supposing that voting occurs sequentially issue by issue. The voting paradox is avoided, but the order of voting becomes crucial and ultimately arbitrary.

3.1.2 Income tax

The archetypal policy instrument for redistribution is a system of taxes and transfers related to pre-tax income. As such, the politics of the progressive income tax has been a leading case in positive analyses of redistribution.

The framework in which the issue is naturally posed is that of the optimal tax literature. We suppose then that people have entirely self-regarding preferences of the form $U(C, L)$, where C denotes the consumption of some composite commodity and L is the amount of labour supply. Individuals differ only in their pre-tax wage w, whose distribution, described by distribution function $F(w)$ on support Ω, is common knowledge but each individual's wage being their own private information. The government announces a tax schedule $T(Y)$ and each individual then chooses their labour supply to maximise $U(Y - T(Y), L)$, where $Y \equiv wL$ denotes gross income. To focus on redistributional issues, we take as given the amount that the government must raise for spending on goods and services; and for simplicity we take that amount to be zero. The problem addressed in the optimal tax literature is then that of characterising the schedule T which maximises some social welfare function or—along the lines discussed in Section 4—that of characterising the set of Pareto efficient schedules. Here our concern is with the schedule that will emerge in political equilibrium; and so, inevitably, with the prior question as to whether there will exist an equilibrium.

Linear income tax

Suppose first that for some reason—perhaps administrative, perhaps the need to eliminate arbitrage opportunities—only linear income taxes are used, those, that is, comprising a constant marginal tax rate t and a poll subsidy α. The policy choice to be made may thus appear two-dimensional, but the government's revenue constraint

$$\alpha = t \int_{\Omega} wL[(1-t)w, \alpha] dF(w) \tag{7}$$

implies a relation between t and α which reduces the policy space to that of the single parameter t. Denoting by $v[(1-t)w, \alpha]$ the indirect utility of a consumer with pre-tax wage w and assuming for simplicity that there are no income effects on labour supply, the welfare that a type-i enjoys at tax rate t is then $\Gamma(t; w_i) \equiv v[(1-t)w_i, t \int wL[(1-t)w]dF(w)]$. Differentiating, using Roy's identity and normalising the marginal utility of income to unity gives

$$\frac{\partial \Gamma(t, w_i)}{\partial t} = \bar{Y} - Y_i - t \int_{\Omega} w^2 L_w^c dF \tag{8}$$

where $\bar{Y}$ denotes mean gross income and L_w^c the compensated effect of the net wage on labour supply. Intuitively, at unchanged behaviour, type-i persons gain more from

the increased poll subsidy than they pay in taxes if their income is below the mean; but the reduction in the labour supply of others induced by the increased marginal tax rate is harmful in reducing the tax base and hence the subsidy.

Consider then the existence of a Condorcet winner. Here the first question is whether $\Gamma(t;w)$ is necessarily single-peaked in t. Itsumi (1974) and Romer (1975) show by example that it is not. A sense of the difficulty is immediate from (8): it would suffice for single-peakedness that $\Gamma(t,w)$ be strictly concave in t, but it is evident that concavity will turn, *inter alia*, on second derivatives of the labour supply function, upon which theory places few restrictions.[30]

FIGURE 4

As discussed above, there may still exist a Condorcet winner even if single-peakedness fails. Roberts (1977) shows that for value restriction to hold in this model, and hence for majority rule to be transitive, it is sufficient that preferences satisfy a condition he calls *hierarchical adherence*: that for any pair of individuals one earns more than the other under all linear tax schedules. More recently, Gans and Smart (1996) show that a sufficient condition for hierarchical adherence is the Mirrlees-Spence single-crossing condition widely used in the context of analysing optimal taxation and screening models more generally: this is the condition that the level curves of $U(C,Y/w)$ in (C,Y)-space are steeper for less able individuals,[31] for which it is sufficient that consumption be normal.[32]

That this condition guarantees the existence of a Condorcet winner is shown in Figure 4. This illustrates the choice between two linear schedules, A and B, represented by T_AT_A and T_BT_B respectively. Suppose that person 1 prefers A to B, as shown. The Spence-Mirrlees condition—that 2's indifference curve through 1's preferred point a be flatter than 1's indifference curve at that point—implies that 2's preferred position on schedule A is northeast of α, at a point like β; but then for 2 to prefer schedule B would require that $IC^{2\prime}$ cut IC^1 at least twice, which the Spence-Mirrlees condition

[30]Nor is it hard to envisage particular circumstances in which single-peakedness may fail to apply. Imagine, for example, an individual who earns less than the average ($y < \bar{y}$) when the tax rate is low, but then finds herself earning more than the average as further tax increases cause the less able to cease work entirely, and who then ceases to work ($y = 0$) as the tax rate rises still further though others do still choose to continue work (so that α continues to rise with t). Recalling (8), for such a person Γ might be increasing at low t, then falling and then rising again.

[31]Note that, confusingly, the Mirrlees-Spence single-crossing property is conceptually quite different from the Gans-Smart notion of single-crossing described in connection with majority rule in subsection 3.1.1 above (which is rather the property that for all $t \neq t^{'}$ there exist a unique $\hat{w}$ such that $\Gamma(t,w) - \Gamma(t^{'},w)$ changes sign at $\hat{w}$): The essence of the Gans-Smart result being described here is that single-crossing in the Spence-Mirrlees sense implies single-crossing in the Gans-Smart sense.

[32]Meltzer and Richard (1981) note that normality of x is sufficient for hierarchical adherence, though not the link with the Mirrlees-Spence condition.

precludes. Thus if 1 prefers A to B so do all those higher in the ability distribution than 1.[33] The schedule most preferred by the person of median ability will muster a majority against any other, and so will be a Condorcet winner. In these circumstances (8) implies that the equilibrium marginal tax rate under majority rule is given by

$$\frac{t}{1-t} = \frac{\bar{Y} - Y_m}{\int_\Omega wL\eta^c dF} \tag{9}$$

where a subscript m relates to the median and $\eta^c[(1-t)\omega]$ denotes the compensated elasticity of labour supply with respect to the net wage $(1-t)w$.

The most immediate property of the equilibrium tax is that it is positive, provided only the distribution of labour income is positively skewed—which we presume to be the case.[34,35] More generally, (9) does not represent a closed form solution for t since elements on the right-hand side themselves depend on it. But inspection suggests that the equilibrium tax system under majority rule will be more progressive (t higher):

- The more skewed the distribution of income. It is natural to take an increase in skewness to be synonymous with an increase in inequality, but this is not necessarily correct: one can have an unambiguous increase in inequality (in the Lorenz sense) combined with a reduction in skewness. Nevertheless many—like Alesina and Rodrik (1994)—interpret skewness as a metaphor for inequality. On this interpretation, greater inequality leads to more intensive attempts at redistribution.

- The less responsive is labour supply. A higher value of η^c in the denominator on the right of (9) points to a lower tax rate on the left; and, more subtly, higher values of t also serve to close the gap between $\bar{Y}$ and Y_m in the numerator on the right.

Clearly, the form of (9) reflects voters' awareness of a fundamental equity-efficiency trade-off in this model, the median voter being constrained in the amount of redistribution she can have go her way by the reduction is the size of the pie available for redistribution as t increases. We shall find an analogous trade-off at work when we analyse the optimal linear income tax in a welfaristic context in Section 4: the

[33]Notice too that this argument shows that the Mirrlees-Spence condition implies that for any linear tax all those with higher ability than 1 will also earn higher income, verifying our claim in the text that it implies hierarchical adherence.

[34]Neal and Rosen (this volume, chapter 7) report evidence on, and possible explanations for, such skewness.

[35]This clearly reflects our assumption that labour supply is independent of lump sum income; it is sufficient for the conclusion to holds more generally that leisure be normal and aggregate uncompensated labour supply increasing in the net wage.

beneficial effects of increasing t to mitigate inequality being offset by efficiency losses due to a labour supply.

Aumann and Kurz (1977) take a somewhat different tack to describe how a majority voting equilibrium might emerge in a linear income tax setting.[36] They assume that voter citizens participate in a multi-person cooperative game with side payments. Voters form into two coalitions—winning and losing ones—which then engage in Nash bargaining to determine the tax rate. The outcome is tempered by the presumption that those in the losing coalition can threaten to destroy their endowment rather than having it confiscated by the majority. The bargaining process ensures that an equilibrium exists, but it is one with a relatively high tax rate.

Non-linear income tax

Far more fundamental than the linear tax case is that in which taxes may be non-linear. This is the natural setting in which to pose some of the most critical issues in the positive analysis of redistribution. Director's Law, for example —the notion that redistribution will run from the extremes to the middle class—is not well-captured by a linear redistributive scheme. Nor can the contrast between universality and means-testing be captured with great richness without allowing for non-linearity: the essence of means-testing being the targeting of benefits on the poor by means of schemes that impose especially high marginal tax rates on them.

It is a weakness of the literature that little progress has been made in characterising the outcome of voting over non-linear tax schedules. This reflects the technical difficulty of the problem. Without some restriction on the set of non-linear schedules over which people vote it is clear that there can in general be no Condorcet winner: for any schedule there will be another, preferred by a majority, in which a previous loser and a winner exploit another winner to the benefit of both.

One approach is to impose essentially arbitrary restrictions on the set of non-linear schedules over which votes are taken. Gans and Smart (1996), for example, consider the case in which attention is confined to equal-revenue tax schedules that can be ranked according to Lorenz dominance. Hemming and Keen (1983) have shown that if one tax schedule Lorenz dominates another and so exhibits more progressivity, they will single cross each other (yet another distinct use of that term). There will be a critical level of income such that those above (below) have lower (higher) net income under the more progressive tax than under the less progressive tax. It is then easily seen that the Spence-Mirrlees single-crossing property implies that among all such tax schedules the Gans-Smart single-crossing holds; and thus that there exists a majority winner.

But the restriction on the set of admissible schedules underlying this result has no deep rationale. An alternative approach is taken by Roell (1996), who confines voting

[36] Inman (1987) provides a particularly accessible account of their analysis.

to those schedules that are most preferred for some voter, perhaps on the grounds that elected citizens cannot commit to pursuing other than self-interested policies. Even in this case the existence of a majority winner does not seem to be assured, though Roell is able to establish existence for the case in which the income elasticity of demand for leisure is zero. One very striking result does emerge, however, from the characterisation of the tax schedules that each voter will selfishly prefer. The voter's only concern with others in this context is as a source of revenue for redistribution to themselves. In considering this problem, of course, each voter must pay attention to the incentive effects of the tax schedule. This, Roell shows, leads each (under plausible conditions) to prefer a negative marginal tax rate—that is, an earnings subsidy—at levels of income below their own. The intuition follow from an observation of Stiglitz (1980) on the shape of the optimal tax schedule in the two-person self-selection model (discussed in Section 4.3): if redistribution runs from poor to rich—perverse in the two-person case, but the relevant comparison here, with each voter regarding all others, including those poorer than themselves, as resources to be exploited—the optimal marginal tax rate on the better-off individual is strictly negative. Loosely speaking, the negative marginal tax rate provides a way of transferring income to the better-off while at the same time inducing a higher level of effort by those better off that will reduce the incentive for the poor to mimic. We return below to the crucial role that the shape of the marginal tax schedule plays in mitigating the incentive to mimic, that is, in inducing incentive compatibility in the choice of a tax schedule which separates individuals according to productivity.

3.1.3 Redistribution and growth

The basic majority voting model has been applied in dynamic settings of endogenous growth models, most prominently by Bertola (1993), Alesina and Rodrik (1994), and, in a framework closer to the above, by Persson and Tabellini (1994a). (See also the overview by Bertola in Chapter 9 below.) The latter consider an overlapping generations model in which the currently young, once they learn their own endowment, vote over a proportional tax on capital income to be imposed next period, with the proceeds being paid to their generation when old as a lump sum subsidy. Each individual then faces essentially the same trade-off as in (8) above.[37] Preference restrictions imposed by Persson and Tabellini (1994a) imply single-peaked preferences, so the equilibrium tax on capital is that preferred by the person with the median endowment. Exactly as above, the tax rate is strictly positive so long as the distribution is positively skewed. Instead of reducing labour supply, however, this tax reduces capital accumulation (with only a substitution effect at work, because the proceeds are redistributed as a lump sum) which in turn reduces the growth rate (because of the beneficial effect

[37] To which equation (7) of Persson and Tabellini (1994a) is essentially identical.

of capital on next period's wage rate).[38] Hence their central conclusion is reached: inequality incites intensified redistribution, which retards growth. Alesina and Rodrik's central point is essentially the same, though it is developed in a model in which redistribution is less explicit: tax revenues finance public infrastructure which boosts wages. They find that unless the policy chosen is that preferred by those who hold only capital, not labour, then policy will not maximise growth.

Further theoretical developments and the sizable empirical literature that has emerged in this area are reviewed by Bénabou (1997) and Bruno et al (1998). Empirically, the suggestion which emerges is that higher levels of initial inequality (especially of land) are associated with, if anything, *lower* growth rates. But this may reflect factors other than redistribution. For example, if economies of scale in mass goods are important in spurring growth—along the lines of Murphy et al (1989)—then inequality may retard growth for quite different reasons. Indeed Bénabou finds that the literature has failed to establish any strong direct effect of redistribution—measured variously by the shares of public expenditures, such as education, health, welfare or transfers, in GDP, or different marginal and average tax rates—on growth, concluding that redistribution is if anything positively related to growth.

Causation could indeed run the other way: from growth to redistribution. Wright (1996) develops a model of majority voting over social insurance schemes in which, with some permanence in policy decision, individuals vote today knowing that even if they have a favorable income draw this period, they (or their descendants) may not in the future. So they trade-off income today against buying insurance tomorrow. This turns on relative risk aversion: Wright shows that if relative risk aversion is less than unity, then faster growth rate will lead to more transfers to the poor and higher taxes on the rich.

3.1.4 Intergenerational redistribution

Redistribution between generations can take many forms, and can run in either direction. Taxes paid by the currently young may help to finance health or residential care for the currently aged; taxes paid by today's workers may finance the education of tomorrow's. Much of the discussion of intergenerational redistribution has been cast, however, in the context of pension provision—'social security', as the term is used in North America (usage being rather wider elsewhere)—though it is recognised that this is a metaphor for intergenerational redistribution more generally. The focus then is on the extent of intergenerational redistribution towards the elderly that one might expect through the adoption of pay-as-you-go (PAYG) pension schemes:

[38] This is not in itself cause for any adverse reflection on the political equilibrium: growth maximisation is generally not an appropriate objective in these models, any more than maximising aggregate output is in the model of labour taxation above.

schemes, that is, under which transfers to the currently old are financed not by the past contributions of the old themselves but rather by transfers from the currently young. This issue suffices to raise the tension between two key forces likely to shape the extent of intergenerational redistribution more widely.

The first—originally emphasised by Aaron (1966) and developed more explicitly by Browning (1975), Townley (1981) and Boadway and Wildasin (1989)—is the incentive that the relatively old have to extract transfers from the relatively young, and the consequent possibility of a political equilibrium in which the social security system is too large in terms of the *ex ante* welfare of each generation but is nevertheless supported *ex post* by a majority of voters. To see how this can arise, consider a simple overlapping generations economy inhabited by selfish people who live for several periods, working when young and retired when old, and a simple pay-as-you-go social security scheme in which the pension in each period is financed by the taxation (at a time-invariant rate) of contemporaneous labour income. Imagine too that (pre-tax) factor prices are exogenous (perhaps tied down in world goods markets) and do not change over time, so that each new generation faces essentially the same fundamentals as its predecessors. Finally, suppose that the social security system is to be decided, once and for all, by a single majority vote. In deciding how to vote, each generation balances the loss it suffers from paying a higher tax rate during whatever remains of its working life against a gain in terms of a higher pension. Then the youngest generation, in looking to maximise its own lifetime utility, effectively seeks to maximise the utility of each generation to come; in that sense, it prefers the scheme which is socially optimal. For each older generation, however, part of the working life has already gone, and so the impact of taxation at such earlier points in the life cycle is ignored. For the retired indeed, the working life is over and all that matters is the impact on pension paid. With the adverse effect of a tax increase during the working life partially ignored, the preference of each older generation is thus tilted toward a higher tax rate than the youngest generation would prefer. So long as the median voter is not to be found amongst the very youngest group, majority voting can thus be expected to lead to a social security scheme larger than is optimal. Thus Aaron shows, for example, that even if $r > n$, so that an unfunded scheme lowers steady state welfare—a PAYG scheme may be adopted in a majority vote.[39]

Notice that even though we have assumed the vote to be once-and-for-all, our as-

[39] An example may help. Suppose that individuals live for three periods, providing some fixed amount of labour in the first period and being retired in the last two. Denoting the tax payable when working by τ and the benefit received when retired by B, the net social security wealth in the first period of life is $-\tau + B\Lambda(r)$, where $\Lambda(x) = (1+r)^{-1} + (1+r)^{-2}$. With population growth at the rate n, the budget constraint for the scheme is that $B\Lambda(n) = \tau$. Since $\Lambda(x)$ is decreasing, the young will support $t > 0$ only if $r < n$, so that the no-tax position is dynamically inefficient. The retired however, look simply to increase B and so (in this simple set up) always prefer $t > 0$. So long as the retired outnumber the workers (requiring that $n(1+n) < 1$), there will be majority in favour of $t > 0$.

sumptions imply that the same scheme would be confirmed if for some reason another once-and-for-all vote had to be taken. What is important to the argument is thus not simply the (in)frequency of voting. Rather the starkness of the conclusion reflects two features of the model. One is the incompleteness of the franchise, with future generations of necessity excluded from the vote: if the untold masses of the unborn generations were able to participate in the vote described above, their interest would align with that of the youngest living generation, and the efficient scheme would be adopted. The other source of the over-provision result is the assumption that it is possible to commit today to tax and benefit rates in the future. For unless the retired are in an absolute majority, they will require political support from the older workers to sustain a PAYG scheme; and those workers will only support the scheme if they themselves believe that succeeding generations will support them in future.

This brings us to the second general key force in intergenerational politics, Tullock's (1997) credibility problem: since workers are typically more numerous than the retired, why do they not renege on any promise they have made—or which the now-old previously made for them—to support the retired? The prospect thus arises of there being too little redistribution towards the old rather than too much. Even if $r < n$, for instance, the possibility that the next generation will renege may prevent the adoption of a Pareto-improving social security scheme.

Early attempts to address this issue simply presumed some ad hoc link between contributions today and benefits tomorrow: see, for example, Hu (1982) and Verbon (1987). More recently, the focus has been on identifying potential devices whereby the credibility problem may be overcome. Sjoblom (1985) envisages an efficient outcome being supported by punishment strategies to penalise generations that fail to support the old of their time; Kotlikoff, Persson and Svenson (1988) describe a reputation mechanism for overcoming such problems. Tabellini (1991) points out that altruism of the young towards the old may also generate political support for payment of pensions by enough of the young to create a winning coalition: in particular, the poor young may support taxes to pay pensions to the elderly, since the bulk of those taxes will be borne by the better-off of their own generation. That in turn raises the possibility that each generation will make inadequate provision for its own old age in the knowledge that tomorrow's young will choose to support them, a samaritan's dilemma of the kind discussed in Section 2. Forced saving may emerge as a means to precommit against excessive redistribution to the old.

3.1.5 Public provision of private goods: Ends against the middle

Many governments spend considerable resources on providing private goods at prices that are substantially below marginal cost, and often zero. Basic health and education services, for example, have many characteristics of private goods yet are often provided free. The implicit redistribution this entails raises two sets of questions. First, can

the public provision of private goods be an efficient way of pursuing distributional objectives? We address this in Section 4. Second, what are the political forces that sustain such redistribution? That is our concern here.

The issue was first addressed by Usher (1977), who considered the likely political support for a system under which the public sector uses a proportional income tax to acquire the entire supply of some private good, which it then makes available to all citizens in uniform quantity. Those with income below the mean benefit financially from the scheme, receiving more as an implicit subsidy on the private good than they pay in taxes. Assuming all to have the same preferences, the right skewness of the income distribution then implies that there will indeed be majority support for the scheme. Against this, however, Usher emphasises the potential role of heterogeneity of preferences in creating political support for private purchasing.[40] Wilson and Katz (1983) extend the approach to the analysis of public subsidies for the purchase of private goods.

These analyses assume that the public sector becomes the sole provider of the private good. In many contexts, however, consumers are left the alternative of purchasing the private good for themselves instead of—or perhaps as a top-up to—the rationed amount on offer from the public sector. Parents may be allowed to take their children out of the state sector and into fee-paying private schools, for example, or to buy private medical insurance. In opting out, however, one cannot opt out of paying the taxes which finance the benefits of those who remain in. Epple and Romano (1996a, 1996b) show that this can have profound effects on the nature of the political equilibrium.

Consider, to fix ideas, the case in which the government will provide some fixed amount of education to each voter, financing this by a proportional tax on their exogenous incomes (the only respect in which they differ), while allowing all to opt out entirely and purchase whatever amount of education they please in the open market. Using the government's budget constraint to eliminate one of its two choice variables—writing the per capita provision level as a function of the tax rate, say—the issue space is one-dimensional. But preferences may fail to be single-peaked over this dimension even though the underlying utility function is perfectly well-behaved. Take an individual with a relatively strong liking for the good at issue. At low tax rates, and hence low levels of provision, she may choose to opt out, in which case further increases in the tax rate are clearly unwelcome; but as the tax rate rises she now opts in, and over some range is likely to support further tax increases.

Failure of single-peakedness precludes the usual easy invocation of the median voter theorem. Recall, however, that the median voter will still be decisive if the single-crossing condition discussed above holds. Epple and Romano (1996a) show that it does

[40] This conflict between the redistributive effect of public provision and the inefficiency of uniform provision is emphasised and explored in a normative context by Weitzman (1977)

indeed hold if the marginal willingness to be taxed in order to pay for public provision does not increase with income.[41] But that marginal willingness might equally well increase with income, in which case the existence of a Condorcet winner becomes problematic. If a majority voting equilibrium does exist, however, it has a striking property: the level of public provision is determined by the balance between a coalition of rich and poor on the one hand and a solid phalanx of middle-income voters on the other. The intuition behind this 'ends against the middle' property—which has evident similarity to Director's Law—is straightforward: the poor would rather have cash than education, and the rich would rather buy their preferred, higher level of education for themselves.

The insight is an appealing and powerful one. Pestieau (1998), for example, arrives at a similar result in pension provision. Indeed it seems likely to emerge from a range of problems in which voting is over the provision of private goods in something like uniform quantity to all: the poor are so poor that they would rather not be forced to pay for the good, while the rich are so rich that they would rather buy their own preferred quantity. In this identity of interests may lie an understanding of the closely related phenomena of Directors' Law and, by the same token, Disraelian Tory democracy.

3.2 Representative democracy

Policy is rarely determined by the entire citizenry meeting in plenary. Rather democracy works principally through the election of officials. Such representative democracy presumably offers advantages over direct democracy in terms of transactions costs, and perhaps too in some of the strategic advantages of delegation. Persson and Tabellini (1994b) argue, for example, that if politicians are unable to commit themselves to pursue policies other than those which they themselves prefer—an issue to which we return below—then one way in which voters may be able to mitigate the time inconsistency problem in the taxation of capital income is by delegating tax-setting to politicians less inclined to tax capital than would be the median voter under direct democracy. Thus one might expect the institution of representative democracy in itself to lead to systematically less redistribution away from capital. It is in the context of representative democracy that political parties have a role to play. Downs (1957) argues that representation, especially by political parties, allows intensities of preference to be registered in ways that direct democracy does not. Political parties can put together a platform which implicitly amounts to vote trading, thereby allowing the collective decision rule to take account of more than simply preference orderings. He argues that this induces economic efficiency in political outcomes, including in those

[41] More precisely, the condition is that the increase in the tax rate that just leaves the consumer indifferent after an increase in the level of per capita public provision on offer decreases with the level of income.

involving redistribution. Our concern here, however, is not to rationalise the prevalence of representative democracy but to explore its implications for distributional policies.

Representative democracies create a role for politicians, whose interests may now have a distinct impact on the outcome. One question that immediately arises is who in society—and how many of them—will emerge as politicians. Almost all of the literature simply assumes from the outset that elections are fought between only two candidates, essentially arbitrarily chosen. Osborne and Slivinski (1996) and Besley and Coate (1997) seek to endogenise the set of candidates by conceiving of each citizen deciding whether to run for office in order to implement their most preferred policy and, perhaps, enjoy some spoils of office. But since the implications for this richer approach for matters of redistribution remain as yet unclear, we here follow the well-trod route of simply supposing there to be two candidates, 'left'(L) and 'right'(R).[42]

The central questions are: How much redistribution would one expect to observe? And how much divergence would one expect between the policies pursued when in office by distinct parties?

The two candidates, we suppose, play Nash in their promises. The winner is the one with a majority. Given this, two broad sets of considerations shape the answers. The first is the degree to which politicians care not only about being in office—and we assume throughout that they derive some 'ego-rents' (that is, some psychic enjoyment, with no resource cost) merely from being elected—but also about the policy that they propose and (or, if defeated, that their opponents) implement. This is important because of its effect on the credibility of promises made to the electorate: politicians known to have their own preferences on policy may find it hard to commit themselves to policies other than those they prefer themselves. The strength of policy preferences thus shapes—but does not entirely determine, since reputational and other devices may also play a role—the extent to which candidates can credibly trim their own position in pursuit of electoral support. The second key consideration is the nature and extent of asymmetric information between voters and candidates: as Downs (1957) emphasised, candidates may be unsure exactly how many votes a particular policy will attract, and voters may not know exactly what policy preferences the candidates hold.

[42]Both Besley and Coate (1997) and Osborne and Slivinski (1996) find that under plurality rule—victory going to the candidate with most votes—under quite weak conditions the equilibrium has only two candidates standing (so to some degree confirming a conjecture known as Duverger's Law). But this provides no support for the usual practice of simply assuming only two candidates: in these models, endogenous candidates cannot commit to pursuing any policy other than their most preferred, so that the equilibria are quite different from those studied in most two-candidate models.

3.2.1 Office-seeking

Suppose first that politicians are entirely apolitical, in the sense that their sole concern is with being elected: they are prepared to, and credibly can, promise any feasible policy in order to get elected. The outcome in the one dimensional case—voting over a linear income tax, for example—is easy to see. Suppose that voters' preferences are known to the candidates and, for simplicity, are single-peaked. Then the same logic as led to the adoption of the preferred policy of the median voter under direct democracy implies that in equilibrium both candidates will offer the policy most preferred by the median voter: this is an instance of Hotelling's Principle of Minimum Differentiation. It has the obvious but important implication that the results for direct democracy above can also be interpreted as applying to a representative democracy in which there are (exogenously) two apolitical candidates. If candidates are uncertain about voters' preferred positions—but have the same distribution over the possibilities (perhaps having access to similar opinion polls or similar experience on doorsteps) then there will again be convergence to a common platform, though to exactly what is less clear.

More novel possibilities arise in voting over multi-dimensional issues. And indeed it is here that representative democracy comes into its own, reducing the dimensionality over which votes must actually be cast. Consider the case of lump-sum redistributive taxes, the extreme form of non-linear taxes. For this case in isolation there is clearly no equilibrium. This reflects the empty core of the redistribution pie problem noted earlier, where each candidate can always find some schedule that a majority will prefer to whatever schedule is offered by his opponent. But individuals may vote not only over the extent of self-interested redistribution, but also over a range of other matters: foreign policy, domestic governance, charisma, and so on. These might in turn be summarised as matters of party loyalty. One such approach to redistribution in representative democracy—developed by Coughlin (1986), and Lindbeck and Weibull (1987, 1993)—is to suppose that party loyalties are private information, though candidates know their distribution in the population: and know their distribution, we shall suppose, conditional on pre-tax/transfer income. Suppose, for example, that individual i will vote for candidate L rather than R iff $U(Y_i + T_{Li}) > U(Y_i + T_{Ri}) + \gamma_i$, where T_{Si} denotes the transfer that party S commits to make to each individual with income Y_i if elected, γ_i denotes i's intrinsic preference for party R (presumably reflecting aspects of R's platform that do not directly affect i's net income—ideology in a broad sense). This leads to a simple model of probabilistic voting, with candidates unsure exactly how individual voters will choose to vote but able to evaluate the impact of their policy announcements on total support received.

Suppose then that the distribution of γ amongst those with pretax income Y_i is given by the distribution function $\Phi(\gamma_i; Y_i)$ (the conditioning on Y being to allow for the possibility that party loyalties may differ systematically across income groups). Each

party, assume, seeks to maximise the expected number of votes it receives,[43] so that L chooses (T_k) to maximise $\Sigma_k N_k \Phi(\gamma_k : Y_k)$, or:

$$\Sigma_k N_k \Phi\Big(U(Y_k + T_{Lk}) - U(Y_k + T_{Rk}); Y_k\Big) \tag{10}$$

subject to the budget constraint $\Sigma_k N_k T_{Lk} = 0$, N_k being the number of voters with pre-tax income Y_k. It is straightforward to show that each party offers the same pattern of transfers (T_k) in equilibrium: since there is nothing they can do to affect party loyalties, each must look only to serving voters' self-interest.[44] The common first-order condition gives

$$\phi(0; Y_k)U'(Y_k + T_{Sk}) = \kappa, \ \ \forall k, \ \ S = L, R \ . \tag{11}$$

where $\phi = \partial\Phi/\partial\gamma$ denotes the density of γ_k. Then an income group k has post-tax/transfer income above the average only if its $\phi(0; Y_k)$ is above the average. This quantity is the proportion of individuals with income Y_k who are indifferent between the parties, in the sense that they would be equally happy voting for either if they offered the same tax schedule; it is thus a measure of the 'waveringness' of group k. The implication of (11) is thus that it is not the groups which show most party loyalty that are receive the most generous transfers, but those who show least loyalty. Indeed the outcome is exactly as if government maximised a social welfare function of form $W \equiv \Sigma\phi(0; Y_k)N_k U_k(Y_k + T_k)$: a utilitarian social welfare function, that is, but with the key difference that groups are weighted not by the number of their members but by their waveringness.

Notice that the equilibrium tax structure implied by (11) is not even necessarily progressive, nor—weaker still—need the transfer even fall with pre-tax income. Concavity of U points towards equalisation, so that if $\phi(0; Y)$ is independent of Y, one obtains complete equalization, as in Section 2.[45] But this effect can be dominated by the tendency to inegalitarianism if the poor are sufficiently determined in their views.[46]

[43] An alternative is to suppose that politicians seek to maximise the probability of their securing a majority. Only in special cases will the two objective functions lead to the same conclusion: Lindbeck and Weibull (1987) show, for example, that (11) below continues to hold in this alternative case iff $\Phi(0 : Y)$ is independent of Y. (See also Grossman and Helpman (1996)). Neither approach is intrinsically preferable to the other, and indeed a more general approach would accommodate both: politicians may care both about being elected and their majority when in office. The modelling choice is essentially arbitrary, but for present purposes nothing critical appears to turn upon it.

[44] That this is the only equilibrium is established in Lindbeck and Weibull (1987, 1993).

[45] Mueller (1989), drawing on Coughlin (1984) also discusses how models in which voter support is uncertain and in which candidates maximise their expected votes, can lead to a unique egalitarian majority equilibrium in the case where voting is over redistribution of a fixed amount of resources.

[46] The ambiguity remains even when incentive effects are admitted and attention confined to linear

One further implication, emphasised by Lindbeck and Weibull (1987) and Dixit and Londregan (1996), is of particular interest. Suppose, for some unexplained reason, that high income persons choose to be affiliated with one party, and low income persons with the other. Then ϕ_k has inverse-U shape over before-tax income. So (11) implies that middle-classes will be especially favoured: in this special case, Director's Law thus emerges again.

3.2.2 The role of ideology

Consider now the case in which politicians care not only about office but also about policy.

One key issue then is whether or not they can commit to adopt policies other than those they most prefer themselves. Suppose first that they can; for there are indeed devices—the desire for re-election, party discipline (retiring politicians penalised if have broken implicit contract)—that may serve such a commitment purposes.

Reverting to the one-dimensional case, in the absence of uncertainty, there will again be convergence of both candidates to the position of the median voter. Whatever their own preferences, both left and right propose and implement, for example, the linear income tax preferred by the median voter: intuitively, it is always preferable to implement a policy proposed by one's opponent than have the opponent implement that same policy. Matters are different, however, if there is uncertainty as to the preferences of the electorate. In this case each party will adopt a platform somewhere between its own preferred policy and the distinct policy offered by the opposition. For suppose, to the contrary, that each offered the same policy in equilibrium. Then each would perceive an advantage from shifting policy towards its preferred point: for since the initial policy maximised the probability of election given the opponent's policy, such a change has no first-order effect on the probability of being elected but since (for at least one candidate) that common platform is away from its preferred policy such a policy shifts raises the politician's welfare if elected. Policy preference on the part of politicians thus sustains a partial divergence of platforms, a point emphasised by Alesina (1988). In the context of probabilistic voting along the lines above, Lindbeck and Weibull (1993) show that[47] each politician announces a policy somewhere between that which maximises $\Sigma_k \phi_k(0)u$ and that which is preferred in their own preferences.

tax schedules. In this case the equilibrium tax rate can be shown, using (8), to be

$$\frac{t}{1-t} = -\frac{cov[\phi_k^*(0), Y_k]}{\int_\Omega \omega l \eta^c dF}$$

where $\phi_k^* \equiv \phi_k / \Sigma_i \phi_i$. From (27) below, the outcome is thus as if the maximisation were of a social welfare function $\Sigma \phi_k(0)\Omega(t, w_k)$; and the marginal tax rate will be positive if, on average, low earners are also waverers.

[47] This is for the case in which the number of policy options is finite.

The ideological component of voters' preferences in the probabilistic voting model above—captured by the exogenous constant γ—plays no real role in shaping the outcome. Recent contributions have given it a more purposive role.

Dixit and Londregan (1998) put more flesh on γ, and allow party platforms to influence it, by supposing that voters differ in the relative weights that they attach to (particular measures of) equality and efficiency in the design of tax policy. Thus parties can influence voter's intentions not only through the effects on voters' own incomes but also by the wider distributional policy they propose. Politicians themselves are assumed to care not only about holding office but also about the distributional position they adopt: party platforms will therefore diverge. In the equilibrium of such a game, it emerges, given the parameterization Dixit and Londregan adopt, that each party offers a tax schedule characterized by a single marginal tax rate (differing across parties in reflection of their differing distributional positions) but with group-specific intercept terms (reflecting the attempt to capture the votes of groups according to the waveringness of their ideological preferences ϕ_i as discussed above, and again in subsection 3.3 below). Not surprisingly, the party more to the left ideologically (that is, putting more weight on equality relative to efficiency) will propose the higher marginal tax rate.

The ideological differences between parties might, however, be on matters quite distinct from redistributional policy: on race or religion, say. Roemer (1998) considers how such matters might affect party platforms on issues of redistribution. There immediately arises the problem of multidimensionality raised earlier: there generally will not exist a Nash equilibrium. Roemer addresses this by proposing a distinct equilibrium concept. Parties are assumed to comprise three types: opportunists, who care only about office; militants, who care only about the platform that is proposed; and reformists, who care about the expected utility of their client group.[48] This leads to the concept of equilibrium as a pair of platforms such that in neither party do all three types agree to a deviation given the policy proposed by the other. In such a setting, Roemer (1998) shows how parties temper their redistributional policy to secure support along the other dimensions of policy: a working-class anti-clerical party, for instance, may offer less than full redistribution to the poor in order to secure support from better-off, anti-clerical voters.

The same equilibrium concept is applied in Roemer (1999) to the pure redistribution problem in which the two competing political parties are identified with the interests of particular income groups. The Left party prefers outcomes most favourable to a particular low-income voter, and a Right party sides with a high-income voter. The policy space is restricted to quadratic tax schedules, which implies using the government budget constraint that two parameters must be chosen, an outcome which would

[48] Thus the objective function in Dixit-Londregan (1998) is a weighted average of the militant and opportunist objectives.

preclude a Condorcet winner in the standard electoral competition for votes. Again, equilibria are defined as those policy platforms such that there are no alternative platforms which would be weakly Pareto preferred by party militants, opportunists and reformers. Then, all such equilibria involve progressive income tax systems.

All this assumes that politicians can commit themselves. Suppose they cannot. If voters know the candidate politicians' types, matters are straightforward. It may be, however, that a politician's type—their preferred progressivity of the tax system, for concretenesss—is their private information. A politician who believes her true type to be unpopular with the voters then has an incentive to disguise that type; similarly, a politician who believes her type to be popular has an incentive to take actions that will credibly signal that type. These issues are considered by Dhami (1997), the framework being one in which voters are uncertain of an incumbent's type—the progressivity of their preferred linear income tax—and term limits mean that the incumbent, if re-elected, cannot run for office again (and so will simply set their preferred tax rate). As one would expect, one possible outcome is a separating equilibrium in which the incumbent finds it worthwhile pursuing a sufficiently extreme policy during their first period in office that their type cannot be mistaken. In the presence of term limits, distributional policy may thus change even though the politician in power does not, a conclusion which finds some echo in the finding of Besley and Case (1995) of greater fiscal volatility in US states with gubernatorial term limits.

3.2.3 The spoils of office

So far we have considered candidates preferences for tax-transfer policies as they affected the welfare of the electorate. But policymakers (politicians/bureaucrats) may have an opportunity to extract resources or rents for themselves, as emphasised in the earlier discussion on kleptocracy. One well-known instance of this is the Leviathan view of government, as posited by Brennan and Buchanan (1980). In the simplest case, the policymaker has as their only objective the extraction of surplus from society. Less starkly, one can envisage policymakers also giving some weight to citizens' well-being, either because of fear of insurrection or a real sense of social justice. Either case is of some interest for our purposes, since the Leviathan tendencies leading a policymaker to divert resources to herself has implications too for extent of redistribution amongst citizens.

To see the forces at work, suppose the policymaker diverts an amount C of the revenue obtained from a linear income tax to herself. The government revenue constraint becomes:

$$\alpha + C = t \int_{\Omega} wL[(1-t)w, \alpha] dF(w) \tag{12}$$

Consider the general case in which the policymaker derives well-being not only from C, but also—a reflection of concerns of the kind just mentioned—from some weighted sum of utilities, $\sum_i \lambda_i u_i$, where $\lambda_i \geq 0$ with strict inequality for some i. The policymaker's problem is then to choose t and C to maximise an objective function given by $\mu(C, \sum_i \lambda_i \Gamma(t, C, w_i))$ where $\mu(\cdot)$ is strictly increasing in both arguments and $\Gamma(t, C, w_i) \equiv v[(1-t)w_i, t\int \omega L[(1-t)w, \alpha]dF(w) - C]$ denotes indirect utility for a household of type i. Assuming there to be no income effects on labour supply, the necessary condition on t is easily seen to reduce to $\sum_i \lambda_i \Gamma_t(t, C, w_i) = 0$. If, following our earlier analysis, there are no income effects on labour supply,[49] Γ_t is independent of C, and the solution for t is exactly as for maximisation of $\sum_i \lambda_i u_i$. Since the policymaker will set $C > 0$, the conclusion is that concern with rent extraction will leave t unchanged but the poll subsidy α lower, so that the concern with rent extraction means a systematic bias towards less progressive taxation than would otherwise be the case.[50] To understand why, it helps to think of the policymaker solving her problem in two stages. In the first, she takes C as given and chooses (t, α) to maximize the social welfare part of her utility function, subject to the revenue constraint implied by using that particular C in (12). In the second, she identifies the full optimum by repeating the exercise for all C. What drives the result is that the constancy of the welfare weights λ and quasi-linearity of household preferences imply that the optimal marginal tax rate at the first stage is unaffected by the revenue constraint (as can be seen from (27) below). Thus the efficient extraction of surplus takes the form of poll tax.

3.3 Interest groups

Many measures with redistributive content appear to be 'particularist' or 'tactical' in the sense that the gains are targeted towards a rather narrowly-defined group. This in turn both reflects and encourages interest groups or lobbies—we use the terms synonymously—to solicit such benefits for their own members. It is to these politics of 'pork-barrel' that we turn in this section.

The connotations of pork barrel and lobbying, of course, are bad. It is as well to remember, however, that lobbies may represent groups whose interests might be widely recognised as more than mere special pleading: lobbies may represent single-parent families, or the aged for example, just as they might represent such easy bogies as farmers, lawyers or the arms industry. In any event, our concern here—as in most of the literature—will not be with the distributional worthiness of the beneficiaries of lobbying activities (which may include, not least, policymakers themselves). Nor

[49] Then $v[(1-t)w, b] = b + h((1-t)w)$ for some increasing $h(\cdot)$.

[50] This ties in with, though with reasoning somewhat different, the Brennan and Buchanan (1980) argument for requiring progressive tax structure as a check on revenue maximisation.

do we consider the formation of lobbies, but rather take their existence as given.[51] Instead we focus on two prior issues on which the literature has focussed:

- Which kinds of lobbies should one expect to be most successful in winning pork barrel for their members? Does it help to be small, for example, and so able to spread the cost of any redistribution over large number of others; or is it better, on the contrary, to have many members and so be able to deliver more votes? Do policymakers reward those with ideological commitments similar to their own, or do they on the contrary attempt to buy support from their natural opponents?

- Will tactical redistribution be delivered in an efficient manner, or will inefficient methods of redistribution be used? At first sight, many forms of redistribution observed in practice appear to be inefficient. A popular example is provided by the Common Agricultural Policy of the European Union. This traditionally benefits farmers by maintaining food prices (to both consumers and producers) above world prices. But exactly the same benefit could be delivered to farmers by simply subsidising their production whilst allowing consumers to trade at world prices: with the benefit to consumers from an increase in consumer surplus more than outweighing the cost of financing the subsidy. Other apparently inefficient policies would include, for example, using entry barriers to support the incomes of professionals (rather than simply paying them a lump sum) and favouring local suppliers in procuring goods for public sector use. The 'Virginia' view—see Tullock (1967) and Buchanan (1980)—is that there is indeed a distinct tendency for particularist redistribution to come in inefficient forms. The 'Chicago' view on the other hand—see Stigler (1982) and Becker (1983)—is that there are incentives tending towards the use of efficient instruments. What appear to be inefficient tools will often turn out, on close inspection, to be second-best efficient, a phenomenon that will recur when we consider redistributive policy instruments in the face of self-selection and commitment problems in Section 4.

Note that there is a link here with the literature on rent-seeking, though the focus there tends to be somewhat different. Rent-seeking refers to the resources and effort devoted to obtaining a share of the rents that are generated by government legislation or regulation—the creation of a monopoly supplier, the protection of products through tariffs, taxes and subsidies, the granting of licenses, and so on. The distortions are typically taken as given and attention is then focused on whether the rents that they create will be dissipated in the competition to enjoy them. Rent-seeking is a risky

[51] Olsen (1965, 1982) provides a wide-ranging analysis of the incentives to form interest groups, and their consequences for redistribution and growth. His analysis is based on the notion that interest groups exist largely to exploit common interests, and the consequent free-rider problems will influence the types of groups that will form and be successful.

business: only some rent-seekers will be successful. They will devote resources until the cost just equals the expected utility of the lottery. Such resource expenditures are pure waste, and may dissipate a substantial part of the rent being competed for. Indeed, if rent-seekers are risk-neutral and symmetric, and if there is free entry, the dissipation will be complete. Here, in contrast, the endogeneity of the distortions is of the essence, and (implicitly) the real resources used in their creation are taken as given (in a sense that will become clearer later).

It is on these two issues that we concentrate. Though the distinction[52] ultimately proves rather strained, it helps organise ideas to distinguish between two approaches to the formal analysis of these issues, differing in the proximate motivation attributed to the lobbies.

The first approach attributes to lobbies only what Grossman-Helpman (1996) call an 'electoral motive': their sole concern is assumed to be with the probability of (re-)electing a policymaker whose interests are favourable to its own. In particular, the lobbies take as given the platforms announced by electoral candidates (and believe they will be implemented if elected).

The question then arises as to what the lobbies have to offer that might affect the electoral outcome. And here it is natural to begin with the case in which all they have to offer is their members' votes. This takes us back to the model of probabilistic voting in the preceding section, from which one can extract a little more in terms of the kinds of groups that are likely to be favoured. Suppose, following Dixit and Londregan (1996), that the utility function in Section 3.2 is of the form $U_i(C) = \frac{\theta_i}{1-e}C^{1-e}$. Solving (11) then gives the share in total consumption of the typical member of group j as

$$\frac{C_j}{\sum N_i y_i} = \frac{\pi_j}{\sum N_i \pi_i} \tag{13}$$

where Dixit and Londregan call $\pi_j \equiv [\phi_j(0)\theta_i]^{1/e}$ a measure of the 'clout' of group j. The lobbies that do best are thus those with the greatest clout. Other things equal, two things tend to place a lobby in a strong position. The first is an ideological indecisiveness, in the sense of a high value of $\phi_i(0)$, the reason being as discussed above. The second is a high degree of venality, in the sense of a relatively large value of θ_j: for θ indicates the weight placed on consumption relative to ideology, and a high value means that a vote is more easily bought.

Three other implications of (13) are of interest. The first is that the poor are inherently advantaged in this game of pork-barrel, irrespective of their clout. To see this, note that if two groups have the same clout then they will end up with the same per capita consumption; and hence the per capita tax will be lower for the poorer group. The reason is simple: the greater marginal utility of income of the poor makes their votes

[52] Which is also made by Rodrik (1995).

cheaper to buy. The second striking implication is that the size of a lobby has no bearing on its clout. Small interest groups are relatively cheap to buy, but they also have few votes to offer: the two effects cancel, leaving the political power of a lobby independent of its size. There is an indirect effect of the size of a lobby on the per capita consumption of its members operating through the denominator of (13), but it is still clout that is critical. Suppose, for example, that all in group i and j have the same per capita income; then an increase in the membership N_j of lobby j combined with a reduction in the size of i will benefit the now-larger group j if and only if it has the greater clout. The third is that both parties offer the same program to the electorate: ideology (which drives voters' intrinsic preferences for the parties) plays no role in the redistribution they offer.

The simplicity of the solution in (13)—in particular, its first-best Pareto-efficiency—reflects the assumption that lump sum taxes are available to the policymaker. Hettich and Winer (1988) examine the choice between distorting taxes in a similar model of probabilistic voting. The characterisation of equilibrium policy is conceptually straightforward: the vote-maximising politician will seek to equate across alternative tax instruments the number of votes expected to be lost per dollar of additional revenue. Though the implied tax structure may be complex, it shares two important properties with that in (13). First, it will be Pareto efficient relative to the instruments available (so long as each lobby group has some positive probability of voting for the party being examined): for it would otherwise be possible to attract more votes from some group without forgoing votes from another. The implication is thus that when only distorting taxes are available the equilibrium tax system will be second best Pareto efficient. Second, all parties again offer the same tax structure.

Inefficient policies may emerge in equilibrium, however, if the government is unable to commit to its future redistribution policies. The argument there is developed by Dixit and Londregan (1995) in the context of support to declining industries. The efficient response to such a decline is likely to involve workers reallocating to higher income occupations, with any compensation deemed appropriate or judicious paid as a lump sum. Once workers move, however, they are liable to find themselves in a new interest group. *Ex post*, policymakers will simply arrange transfers to that group in line with (13) above. Thus any private income gains associated with moving to an interest group with higher income may ultimately be taxed so heavily that the net gain falls short of the cost of moving; then forward-looking voters will not make the move, but instead remain in the declining sector and take such transfers as (13) generates. Suppose, to take an extreme case, that the efficiency-enhancing action requires moving between two groups that have the same clout; then unless the government is able to commit to distinguish *ex post* between those who moved and those who did not, those who move will experience no private gain from doing so. The economy is thus locked into an inefficient policy by the inability of government to commit itself to allow those who move to enjoy sufficient of the efficiency gains they create in doing so. Note though

that this problem is not particular to a world of pork-barrel. Exactly the same time consistency problem arises whenever governments are unable to commit themselves not to redistribute away the private benefits that individuals are able to reap from decisions that enhance overall efficiency: such as investments in education, an example we shall return to in Section 4.

The lobby groups envisaged in the discussion so far have been very passive: all they do is vote. One generally thinks of lobbies as seeking to exercise power in other ways too. A more proactive way in which they might pursue their electoral motive is by contributing to the campaign funds of political parties. For by sending costly messages to rationally uninformed voters—that is, who would otherwise not find the cost of informing themselves worthwhile in terms of the gain from any conceivable change in their voting behaviour—lobbies may affect the distribution of electoral outcomes. It may be, for example, that they are able to provide (perhaps misleading) information on the consequences of policy choices that voters would otherwise not find it worthwhile obtaining for themselves or their contributions themselves convey information to voters as to the magnitude of the income transfers at stake in choosing between alternative policies (Pyne (1997)).

This potential use of campaign contributions to affect election outcomes is at the heart of an influential approach to the analysis of lobbies developed by Magee et al (1989). Suppose there are just two parties and two lobbies, with lobby i contributing, if at all, only to party[53] i, $i = 1, 2$. The probability of party 1 being elected is taken to be some function $\pi(c_1, c_2, \theta_1, \theta_2)$ of the campaign contributions c_i of the two lobbies and the actions θ_i of the two parties. The function π is not derived from first principles, but instead assumed to satisfy plausible restrictions. In particular, π is increasing in c_1 and decreasing in c_2. Politicians choose policies θ and are concerned only to maximise the probability of winning office. Lobby i derives utility $U_i(\theta) - c_i$ from policies enacted, and so chooses its contributions to maximise expected utility $\pi U_i(\theta_1) + (1 - \pi)U_i(\theta_2) - c_i$. Lobbies play Nash against one another, and in doing so take declared policies θ_i as given; that is, contributions are determined after platforms. Parties play Nash against each other, as first movers relative to the lobbies. In the basic formulation of the model, each party recognises that its own policy may affect the contributions of its own lobby, but takes as given the contributions of the other.

The principal prediction of this framework is that the presence and activities of the lobbies may well involve the parties offering distinct policies even though they have no ideological commitments; and, most important for present purpose, they can be expected to sacrifice efficiency to the particularist interests of the lobbies. Suppose, for example, that there is some policy θ^* which maximises the sum of utilities, so that

[53] This can indeed be shown to be the case, under weak restrictions: the only return to contributions is changing the probability of election, so that contributing to both is merely to some degree self-defeating. In a wider context, of course, contributing to both may be optimal if, for example, past contributions, buy access to politicians in office.

a fully efficient solution requires both parties to commit to θ^* and both to make zero contributions. And suppose too, to bring out the distinctive consequences of lobbying, that if contributions were fixed at zero each party would find it optimal to offer the efficient policy, implying that (taking π to be differentiable, for simplicity)

$$\frac{\partial \pi(0,0,\theta^*,\theta^*)}{\partial \theta_1} = 0 \tag{14}$$

In these circumstances it is clear that the efficient polices cannot be sustained in the equilibrium of the model above. For each party would then find it optimal to deviate by offering a policy close to θ^* but more attractive to its own lobby: the change in policy itself will have no impact on the probability of election, from the property just described, but it will induce positive contributions from lobby 1—which now finds it worthwhile to seek the election of that party rather than its opponent—and thus the probability of election increases. More generally, any tendency towards reduced political support consequent upon deviating from an efficient policy may be more than compensated for by the political support that can be bought from the increased contributions of the favoured lobby.

This presumption of inefficiency, however, depends crucially on the assumption that each party takes the contributions of the other party's lobby as given in making its decisions. Clark and Thomas (1995) show that once this impact is taken into account the efficient solution will indeed be sustained in equilibrium. Intuitively, any departure from efficiency must damage the other lobby even more than it benefits the favoured lobby, and thus the other lobby has a greater incentive to campaign against the party proposing the deviation than its supporters have to lobby for it. Thus efficiency comes to predominate despite the willingness of politicians to serve the interests of particularist groups.

This brings us to the second broad approach to the analysis of interest groups: that which sees them as deliberately attempting to affect the policy of a government in office. Elections and electoral concerns are relegated to the background in modelling this influence motive. One interpretation of these models is that they are reduced forms in which the appearance of interest groups ultimately reflects their impact on electoral outcomes. A more coherent rationale would rest on a view that elections are ultimately bought, a view put forcefully by Becker (1983), who rationalises a neglect of voting on the grounds that 'voter preferences are frequently not a crucial *independent* force in political behavior [but] can be manipulated and created through the information and misinformation provided by interested pressure groups.' (p.392). This may be more plausible for relatively narrowly defined particularist redistribution rather than programmatic redistribution. In any event, the key feature of these models is that interests seek to affect the distribution over policy outcomes other than by affecting electoral outcomes.

One simple way of modelling this is to posit some essentially *ad hoc* influence function by which resources spent by lobbies—on advertising, political contributions, or simple bribes—tilt policy in their direction. From such a framework Becker (1983), in a pioneering contribution, developed two influential sets of arguments.

First, he too arrived at a presumption—albeit a carefully qualified one—that efficient methods of redistribution (and its financing) will tend to be used. The reason is similar to, though distinct from, those that emerge from the electoral motive stories: beneficiary interest groups can expect to encounter less political resistance if their handouts are financed by an efficient tax structure, and if the amount that must be raised is reduced by distributing the benefits in efficient ways. An important qualification to this argument, as emphasised by Becker, is that the total amount of support is taken as given; we return to this below.

The second implication is that it is advantageous for a lobby group to have relatively few members: for then the tax rate levied on the rest of society will be relatively low (as will the associated deadweight loss), implying, one might suppose, relatively low political resistance.[54] This is in stark contrast, of course, to the implication of the model of probabilistic voting above that size is of no importance. The advantage that the poor enjoy of being cheap to support is no longer offset by the disadvantage of having relatively few votes at their disposal. This conclusion does indeed seem consistent with the casual observation of the support often enjoyed by such relatively small groups as farmers in the developed countries, workers in declining industries. In practice, however, it is not entirely clear that this is independent of electoral consideration, since such groups may often be large relative to voting constituencies even though small relative to the population as a whole. The role of smallness too will be returned to later.

An alternative approach to capturing a direct influence of lobbies on policy is to suppose that their interests enter into the objective function of policymakers with a particularly heavy weight. Like the influence function, this 'political support' function is obviously something of black box. The common agency approach to the analysis of interest group pressures—developed by Grossman and Helpman (1994), and further refined and applied by, in particular, Dixit (1996) and Dixit, Grossman and Helpman (1997)—takes as its point of departure precisely the attempt to model explicitly the way in which interest groups may influence policy. The essence is to see the policy maker as a single agent responding to, in the interest groups, many principals.

Consider for example the lobbying by firms for publicly financed subsidies on their inputs. Suppose there are M sectors in the economy, with the typical sector i comprising N_i individuals each with access to a technology $f_i(x)$ for producing good i from an

[54] Olsen (1965) argued that size was a disadvantage in interest group formation for another reason. The fewer the members, the easier it is to overcome free-rider problems that inevitably accompany interest groups lobbying for a common interest.

imported input x and some fixed factor owned by that individual. Prices of all goods are fixed on world markets, and normalised at unity. Purchases of sector i are subsidised at rate s_i, and these subsidy rates are potentially the object of lobbying as are, for expositional purposes, lump sum subsidies a_i payable to members of each sector. Profits of the typical member of sector i are then $\pi_i(s_i) \equiv \max_x\{f_i(x) - (1 - s_i)x\}$, giving total income

$$Y_i \equiv \pi_i(s_i) + a_i - c_i(\mathbf{s}, \mathbf{a}) - \frac{1}{N}\{r(\mathbf{s}) + \sum_{j=1}^{M} N_j a_j\} \tag{15}$$

where $c_i(\mathbf{s}, \mathbf{a})$ denotes the contribution that i makes to the policymaker (whether as campaign contribution or other inducement, legal or not)—which, note, are conditional on the policy pursued—whilst $r \equiv \sum N_j s_j Y_j(s_j)$ denotes total revenue (campaign contributions, it is assumed, being necessarily retained by the policymaker), $N \equiv \sum N_j$ denotes the total population and it is assumed that any revenue requirement is raised by a uniform poll tax. The utility of individual i is then $U(Y_i)$, with U concave.

The policymaker cares both about the contributions received and the sum of utilities, with linear preferences between the two described by the objective function

$$G(\mathbf{s}, \mathbf{a}) \equiv \sum_{i=1}^{M} c_i(\mathbf{s}, \mathbf{a}) + \beta \sum_{i=1}^{M} N_i U(Y_i) \tag{16}$$

We assume that the policymaker values contributions more than citizens' well-being, in the sense that she would prefer to receive a dollar in contributions than have a dollar given to any citizen: thus we confine attention to circumstances in which

$$1 \geq \beta U'(Y_i), \quad \forall i \tag{17}$$

The interest groups choose their contribution schedules so as to maximise $U(Y)$.

For the moment, suppose lump sum transfers are impossible, so that $a_k = 0, \forall k$. The first-order condition for the government's choice of s_k is then

$$\sum_{i \in L} N_i \frac{\partial c_i}{\partial s_k} + \beta \sum_{i=1}^{M} N_i U'(Y_i) \left(\delta_{ik} x_k - \frac{\partial c_i}{\partial s_k} - \frac{1}{N} \frac{\partial r}{\partial s_k} \right) = 0, \quad \forall k \tag{18}$$

where L denotes the set of sectors organised into lobbies (which we take to be given exogenously, and unchanging), δ_{ik} is the Kronecker delta and use has been made of the relation $\pi_k'(x_k) = x_k$. To characterise the optimal choice of schedule by the lobbies, note that lobby i will set its schedule so that the policy choice it induces maximises

$G + U(Y_i)$: for since the contribution schedule effectively enables side payments to be made between i and the policymaker, it would otherwise be possible to make both better off by choosing differently. And since we have just seen that policy is chosen to maximise G, it must be therefore be the case that it maximises the utility of each active lobby. Thus

$$\delta_{ik} x_k - \frac{\partial c_i}{\partial s_k} - \frac{1}{N}\frac{\partial r}{\partial s_k} = 0, \quad \forall k \in L \tag{19}$$

Combining (18) and (19) gives[55]

$$x_k \left(\gamma_k - \alpha_L + \beta(1-\gamma_k)U'(Y_k) - \beta \sum_{i \notin L} \alpha_i U'(Y_i) \right) = s_k x_k' \left(\alpha_L + \beta \sum_{i \notin L} \alpha_i U'(Y_i) \right) \quad \forall k \tag{20}$$

where γ_k takes the value unity (zero) if sector k is (not) organised as a lobby, $\alpha_i \equiv N_i/N$ and $\alpha_L \equiv \sum_{i \in L} \alpha_i$ denotes the proportion of the population in organised sectors. Also denoting (minus) the elasticity of demand for the input by E_k, one finds from (20) the equilibrium subsidy to sector k to be given by

$$\frac{s_k}{1+s_k} = \frac{1}{E_k} \left(\frac{1 - \alpha_L - \beta \sum_{i \notin L} \alpha_i U'(Y_i)}{\alpha_L + \beta \sum_{i \notin L} \alpha_i Y'(Y_i)} \right) \tag{21}$$

if k is organised and

$$\frac{s_k}{1+s_k} = \frac{1}{E_k} \left(\frac{\beta(U'(Y_k) - \sum_{i \notin L} \alpha_i U'(Y_i)) - \alpha_L}{\alpha_L + \beta \sum_{i \notin L} \alpha_i U'(Y_i)} \right) \tag{22}$$

if it is not.

An immediate implication of (21), given the assumption in (17), is that all organised groups receive a positive subsidy. The position of the unorganised is less clear-cut: some of them may also receive subsidies, their marginal utility of income being so high that the policy-maker's interest in their well-being justifies favouring them. It is sufficient, however, for all unorganised groups to find their input being taxed in order to subsidise the organised groups, that all such groups have the same marginal utility of income.

[55] Condition (20) below is exactly the necessary condition for maximising an objective function of the form $\sum_i \sigma U(Y_i)$, where $\sigma = \gamma_i / U'(Y_i^*) + (1-\gamma_i)\beta$, Y^* denotes the value of income at the optimum and γ_i is as defined below. The outcome is thus as if the policymaker sought to maximise a political support function in which (by our assumption that $\beta\lambda < 1$) those organised in lobbies receive especially high weight.

Within the set of organised groups the only characteristic that affects the size of the subsidy is the elasticity of demand: the lower this is, the higher is the subsidy. For a lower elasticity means a lower deadweight loss, and hence a greater welfare gain for both the group concerned and, thereby, the policymaker.

The size of an interest group, on the other hand, has no impact on the subsidy that it receives in equilibrium. What matters instead is the size of all active lobbies taken together, α_L. The intuition, it seems, is that since all particularist benefits are ultimately paid for by the unorganised, and since the policymaker weights all individuals equally—in terms of both their welfare and their cash contributions—so it is simply the collective size of the organised groups that determines the rate at which each is subsidised.

The final welfare of any interest group i depends, however, not only on the subsidy rate s_i that it is able to attract but also on the resources $c_i(\mathbf{s}, \mathbf{a})$ it spends in the process. This in turn is tied down by the condition that the payoff to the policymaker be the same as it could obtain by dealing with all active lobbies except the ith: the payoff can be no less, otherwise the policymaker would indeed deal only with those others; and it can be no more because otherwise the interest group could change its contribution schedule so as to leave the marginal incentives facing the policymaker unchanged but the level of contributions lower. Thus the welfare of any interest group will generally depend on the extent and nature of other organised interest groups. One might expect, in particular, the final benefit to any lobby to be smaller—and the net gain to the policymaker greater—the more active groups there are. For suppose, at one extreme, that only one lobby is active in the subsidy game above. Then since in the absence of that lobby all subsidies would be zero, the payoff to the policymaker must be just as in the no-intervention case. Thus the policymaker derives no benefit from the activities of the interest group, with the contribution received exactly offset by the welfare losses of the non-organised groups. All the gains go to the lobby. At the other extreme, if all groups are organised then (21) and (22) imply that all subsidies are zero in equilibrium. If any one group were to be eliminated, however, it would find itself facing a positive tax in order to finance the others. Thus each group will make some contribution simply to avoid being made the victim of all others: all citizens—organised and not—would be better off it the organised could commit not to make any contribution, with all the rents created by their ability to do so being captured by the policymaker.

This reminds us that the analysis takes as given—as indeed we have done throughout—the existence of organised interest groups. One might naturally think of taking this game back one stage further to consider entry decisions by lobbies, though it seems this has not yet been done in this framework.

Consider next the issue of whether one would expect inefficient policies to be used in a world of the kind just described if efficient policies are technically feasible. This can

be addressed by now allowing the lump sum transfers **a** to be deployed. Proceeding as above, the condition analogous to (20) reduces to

$$\gamma_k - \alpha_L + \beta(1-\gamma_k)U'(Y_k) - \beta\sum_{i\notin L}\alpha_i U'(Y_i) = 0 \quad \forall k \tag{23}$$

which, combined with (21) and (22), implies that $s_k = 0$, $\forall k$. Thus only the efficient policy is used in equilibrium. This does not mean, however, that the availability of efficient instruments is Pareto-improving. Organised interest groups may prefer that the policymaker be restricted to using the inefficient subsidy system. Take, for example, the extreme case in which all groups are organised, so that each must in effect pay protection money in order not to be exploited by the others. Then since lump sum taxes provide a more efficient means of expropriation, groups may well find themselves having to pay more protection money as a consequence. The same effect may be at work when not all are organised.

More strikingly still, one can show that at least one unorganised group is certain to prefer an equilibrium in which only the inefficient instrument is available (and used) to one in which efficient lump sum transfers can be made.[56] The intuition is essentially the same as just given for the lobbies, the only difference being that it is sure to apply for at least one group: a restriction to inefficient policies can reduce vulnerability to expropriation.

There is indeed a general principle here: when policy-making is not entirely benevolent, a restriction to the use of inefficient instruments may be socially beneficial. An early instance of this, for reasons quite distinct from those in the model of common agency above, was developed by Rodrik (1986).[57] The argument there is that since free-rider problems are likely to impede the formation of lobbies to seek tariffs (whose benefits are a public good to all firms in the protected sector) more effectively than they impede the formation of lobbies to seek firm-specific subsidies (whose benefits are akin to a private good), a commitment to deploy only tariffs may render the economy less vulnerable to special-pleading. The same principle is present too in the prescription of Brennan and Buchanan (1980) that the citizenry may wish to restrict a revenue-maximising Leviathan to the use of inefficient instruments as a means of limiting the extent to which revenue can be extracted from them: one might, for example, require commodity taxes to be uniform even though there would be efficiency gains from appropriate differentiation. Indeed Brennan and Buchanan would on these grounds

[56] To see this note first that (23) implies that $\beta U'(Y_i^{*A}) = 1$ for all $i \notin L$, where the superscript $*A$ refers to an optimum when lump sum transfers a are available. (This follows on setting $\gamma_k = 1$ in (23) to find $1 - \alpha_L - \beta\sum_{i\notin L}\alpha_i U'(Y_i^{*S})$ and using this to evaluate (23) for $\gamma_k = 0$). Since it is easily seen from (21) that $s_k > 0$ for some k implies the existence of some $i \notin L$ such that $\beta U'(Y^{*S}) < 1$ (the superscript S indicating the regime in which only the distortionary subsidies s are available), there thus exists some $i \notin L$ such that $U'(Y^{*L}) > U'(U^{*S})$, and the result then follows from concavity.

[57] See also Wilson (1990).

require Leviathan to impose a tax schedule that is progressive: for it is well-known that such a schedule is (which cannot have a zero marginal tax rate at the top) cannot always be revenue-maximising.

If one is to explain the apparently widespread use of inefficient means for tactical redistribution in terms of the benefits that a commitment to the use of such instruments conveys on the otherwise vulnerable, some explanation of how such commitment is achieved in practice is needed. One might perhaps rationalise non-discrimination rules in tax legislation as to some degree serving to commit against the use of firm-specific measures of the kind that may be most vulnerable to lobbying. Beyond that, however, it seems one must look to extra-legal commitment devices. And here it is unclear, for example what mechanisms are at work to penalise politicians who replace inefficient instruments with efficient.

It remains to consider one further view, not captured in any of the models described so far but apparently widely held, as to why inefficient policies may be pursued: because the losers find their losses from inefficient policies opaque and harder to perceive than they would under a system of lump sum transfers. A formal treatment of such 'optimal obfuscation'[58] is provided by Coate and Morris (1995). Voters, it is assumed, are uncertain about both the type of politicians' seeking their vote—whether they are 'good' in the sense of caring only for the citizens' well-being or 'bad' in the sense of having a political support function that reflects the interests of some lobby—and the social return on some public project—which is certain, however, to benefit the special interests. Rather than simply pay lump sum transfers to the interest group and thereby reveal their type, bad politicians, it is shown, may implement the particularist project even when they (though not the voters) know it to be socially undesirable.

The notion of optimal obfuscation is reminiscent of, though distinct from, the argument of Becker (1983) and Stigler (1982) touched on above, also apparently widely held, that small lobbies tend to be successful because the cost of the support is spread over so many people that the cost to each is so small—perhaps even unnoticed—as not to warrant any action. Both raise the same counterargument (powerfully argued by Breton (1993): exploitation of an ill-informed majority by a minority creates an opening for a political entrepreneur to make the majority aware of the position and reap the reward of their electoral support. This then raises a range of issues concerning the ability of political entrants to discover the underlying truth—the true social return to the project, in the Coate-Morris model—and credibly convey it to the electorate, who will be aware that the entrant has an incentive to misrepresent the true position. It is thus by no means obvious that the entrepreneurial flair of politicians can be relied on to eliminate inefficient redistributive policies in democracies. Indeed if they could, the problem would become that of explaining why one does not observe more oppressive redistribution away from minorities.

[58] The term originates, it seems, with Magee et al (1989).

4 Constraints on redistribution

In an ideal world, policymakers would have the luxury of choosing among allocations along the economy's first-best utility possibilities frontier (UPF). The first-best UPF represents the Pareto-efficient utility combinations achievable when the only constraints are the resources available and the technologies for converting resources into goods and services. In a first-best world, the decentralization of economic decisions to competitive markets has well-known desirable characteristics, summarized in the *Two Fundamental Theorems of Welfare Economics* (Arrow (1951b)). The First Theorem states that competitive markets are Pareto-efficient regardless of the pattern of ownership of the economy's resources. The Second, more relevant for our purposes, says that, given convex preferences and technology, any Pareto-efficient allocation of resources can be achieved by a suitable reallocation of endowments (that is, lump sum income) among households in the competitive economy. The redistribution problem facing a welfarist policymaker is then to choose a pattern of lump-sum transfers of income so as to achieve the desired point on the UPF.

The choice among various utility combinations depends upon the precise form of the policymaker's social welfare function. For the purposes of this section, we restrict ourselves to considering Pareto-efficient outcomes, those chosen from the relevant UPF. The rationale for this is based on the role of economist as scientific policy advisor rather than as an analyser of actual political decision-making, the latter hat having been worn in the preceding section. We focus on the considerations that determine the alternative utility allocations from which the policymaker might choose, eschewing any attempt to evaluate one versus the other using interpersonal utility comparisons. The analysis will not be completely value-neutral, however, for we typically take for granted both the welfaristic perspective and the Pareto principle. And, typically we illustrate our points using relatively mild restrictions on the ordering of alternatives. In particular, we shall usually suppose that the policymaker's objective function is quasi-concave in individual utilities, an assumption that allows us to concentrate on that portion of the UPF which involves redistribution from the better-off to the less well off. Recall that a quasi-concave social welfare function encompasses degrees of aversion to inequality ranging from zero (utilitarianism) to infinite (maximin).

4.1 First-best redistribution

As a benchmark, it is useful to consider how optimal redistribution policies can vary in an ideal first-best world according to context. Consider first the case in which household incomes are fixed but can vary from one household to another. Household utility depends only on income, which can be used to purchase a composite commodity. If all households had identical strictly concave utility functions and if the policymaker had

any quasi-concave social welfare function, then—essentially for the reason discussed at the outset of Section 2—optimal redistributive policy would imply full egalitarianism regardless of the degree of inequality aversion.

Next, suppose that utility functions differ among households such that some households receive more utility from a given level of income than others, and also have a higher marginal utility at any income level. Some persons, say the disabled, are less efficient 'utility generators' than others. Thus, if person A is a better utility generator than person B, $U_A(Y) > U_B(Y)$ and $U'_A(Y) > U'_B(Y)$ for all incomes Y. In this case, as Sen (1973) observed, optimal redistributive policy depends critically upon the degree of aversion to utility inequality. Under a utilitarian social objective function, all that matters is the sum of utilities, so more income goes to the more efficient utility generators. The result is an unequal distribution of both incomes and utilities. At the other extreme, under a maxi-min social objective function, the point of equal utilities on the UPF will be chosen, with more income going to the less efficient utility generators. For less extreme values for the aversion to inequality, some utility inequality will be tolerated alongside some income inequality, though less than under utilitarianism.

For future purposes, one feature of these outcomes should be noted. To implement the optimal redistributive scheme, policymakers must know which utility function is attached to each household. If not, households cannot be relied upon to report truthfully on their utility functions. Any attempt to invoke an unequal distribution of incomes will be incentive-incompatible: under utilitarianism, persons will all prefer to tell the policymaker that they are highly efficient utility generators, and vice versa for maxi-min (Dasgupta and Hammond (1980)). Given this inability to identify utility functions with individuals, the government can do no better than to equalise incomes, a policy which had been proposed by Lerner (1944) because it maximized expected social welfare. The requirement that income redistribution policies be incentive-compatible when governments have imperfect information is the key limit to redistribution that we focus on later in this section.

Once income is allowed to be variable, matters become murkier still. Suppose, following the standard example used in the optimal redistributive taxation literature (Mirrlees (1971)), that household incomes come from variable labour supplies, with different households able to generate different amounts of income per unit of labour supplied. Let utility functions be the same for all households, with leisure and consumption both normal goods. In this case, the maxi-min objective function still generates equal utilities in the first best. Since their wage rate is higher, higher ability persons supply more labour and earn more income, but since they must reach the same utility level as lower ability persons, they also consume more. The amount of (lump-sum) taxes they pay will be higher, but their average tax rate (taxes divided by income) can be higher or lower depending on the form of the utility function. That

is, the tax system need not be progressive (Sadka (1976)).[59] This is rather surprising, given that the maxi-min social objective function puts extremely high weight on egalitarianism.

The opposite extreme of utilitarianism is equally surprising. As Mirrlees (1974) showed, in this case (and so long as leisure is a normal good) higher ability persons end up with *lower* utility levels than lower ability persons. The reason is that because they are more efficient at producing output, they are required to supply much more labour. The tax system is in a sense highly redistributive, completely reversing the rankings of utility.[60] One the other hand, it may or may not be progressive depending again on the curvature of indifference curves.[61]

The upshot of this discussion is that, even under ideal first-best conditions, there is no presumption about how progressive the redistributive lump-sum tax structure should be, or even if it should be progressive at all. Traditionally, most analyses of optimal redistributive policy have presumed that distortionary taxation must be used. Thus, the process of redistribution induced inefficiencies, forcing redistributionally

[59]Formally, in the maxi-min optimum with identical preferences, all individuals achieve the same utility level: $U(C_i, L_i) = \bar{U} \quad \forall i$. By utility maximization, individual's i's consumption-labour choice satisfies $w_i = -U_L/U_C$. The indifference curve associated with the equilibrium utility level $\bar{U}$ implicitly defines a relationship between C and L, say, $C = g(L)$. Then, by differentiation of the utility function along an indifference curve, we obtain $w_i = -U_L(C, L)/U_C(C, L) = g'(L)$. The average tax rate for person i is given by

$$T_i/Y_i = \frac{w_i L_i - C_i}{w_i L_i} = 1 - \frac{g(L_i)}{g'(L_i)L_i} = 1 - \frac{1}{E(L_i)}$$

where T_i is a lump-sum tax and $E(L_i)$ is the elasticity of the indifference curve. Thus, the tax is progressive iff $E(L)$ is increasing along the indifference curve, which Sadka shows will be the case if the elasticity of consumption of leisure for consumption is less than or equal to one.

[60]The proof is as follows. Let $C(w)$ and $L(w)$ be consumption and labour supply for a type-w individual. Then the change in utility as skills increase is given by $dU/dw = U_C C_W + U_L L_W$ where C_w and L_w are uncompensated price derivatives. In a utilitarian optimum $U_C = k$ for all w, and $U_L = -wU_C$ by consumer maximization. Totally differentiating these two conditions and solving, one obtains:

$$C_w = \frac{U_C U_{CL}}{D}; \qquad L_w = -\frac{U_C U_{CC}}{D}$$

where $D = U_{CC}U_{LL} - U_{CL}^2 > 0$ by the second-order conditions. Substituting these back into the expression for dU/dw, one obtains:

$$\frac{dU}{dw} = \left(\frac{U_C U_{CL} - U_L U_{CC}}{D}\right) U_C$$

Since the numerator is negative if leisure is normal, $dU/dw < 0$.

[61]Using techniques similar to above, Sadka (1976) shows that the tax will be progressive if the elasticity of the curve along which the marginal utility of income is constant is less than or equal to one; but if that elasticity is large enough the tax will be progressive. The intuitive interpretation of this elasticity is not apparent.

minded policymakers to choose from points along a second-best, or nth-best, UPF. The expectation was then that, whatever the form of the social objective function, the progressivity of the tax system will be even less that it would have been under first-best redistributive taxation.

But why must distortionary taxation be imposed? Recent literature has suggested that rather than being imposed externally, distortionary taxes are optimal instruments in a second-best world in the sense that they enlarge the utility possibilities available to the policymaker. This literature stresses that what prevents the government from redistributing along the first-best UPF—equivalently, what causes a violation of the Second Theorem of Welfare Economics—is imperfect information. The government cannot directly discern the better-off from the less well-off, so must induce individuals to reveal their true types through their behaviour. According to this view, to which we turn next, the choice of redistributive tax is a problem in mechanism design. The use of distorting taxes turns out to be an 'efficient' way to induce truthful revelation.[62]

4.2 Imperfect Information as a Limit to Redistribution

The well-being of individuals depends on the set of circumstances facing them—their preferences, the state of their health, the effort they exert, their attitude toward risk, their productivity, and so on. Some of these are likely to be private information to the individual, and this inhibits the government from pursuing its welfaristic objectives. It does not preclude redistribution entirely because the government may be able to observe behavioural and other characteristics of individuals that are correlated with these underlying circumstances. But, imperfect information restricts the utility possibilities, or the efficiency-equity trade-off, available to the government.

The standard way to illustrate this uses the optimal income tax framework for redistribution due to Mirrlees (1971).[63] Differences among households are confined to one characteristic, their ability, which, given the linear production technology and the assumption of perfect substitutability among different ability-types of labour, can be normalized to be the household's wage rate w. For purposes of exposition, we assume a discrete distribution, following Guesnerie and Seade (1982), Stern (1982) and Stiglitz (1982). In fact, we can often simplify the distribution to one with only two ability-types with wage rates $w_2 > w_1$. Though the government cannot observe individual wage rates, it can observe incomes, $Y \equiv wL$. Given an income tax function $T(Y)$, it then also knows consumption $C = Y - T(Y)$. Through its choice of an income

[62]The importance of an imperfectly informed government was recognized by Vickrey (1945) and Van Graaff (1957), and formalized by Mirrlees (1971, 1974). The role of distorting taxation in such a setting out carefully in Nichols and Zeckhauser (1982) and analyzed as a problem of mechanism design by by Roberts (1984).

[63]While Mirrlees solved the optimal income tax problem and drew attention to its implications, Vickrey (1945) had actually set down the formal problem facing the imperfectly-informed government.

tax structure, the government is able to offer any relationship between consumption and income, including highly non-linear ones, subject to its budget constraint being satisfied.

Since the government cannot observe each person's ability, it cannot make a person's tax payment contingent on their ability-type. It can only make it contingent on their income, which can be manipulated by the taxpayer. The redistribution problem is thus a classic revelation mechanism design problem. Taxpayers will only be induced to choose the (C, Y) combination intended for them—that is, to reveal their ability-type—if they prefer that (C, Y) over those intended for other ability-types. This is the source of an *incentive compatibility* or *self-selection constraint* that limits the extent of redistribution possible. Intuitively, higher-ability persons can earn a given income with less effort than can lower-ability persons. As the government attempts to redistribute income from better-off to worse-off persons using income as the indicator, at some point before income equality is reached the higher-ability persons will prefer the income-consumption bundle intended for the lower-ability persons over that intended for themselves; that is, they will be tempted to *mimic* lower-ability persons. At that point, the self-selection constraint binds and further redistribution becomes impossible. In principle, a self-selection constraint must be satisfied between all pairs of ability-types.

More formally, recall the definition of the utility function for a person of type i: $V^i(C, Y) \equiv U(C, Y/w_i)$. As mentioned, indifference curves in $C, Y-$space are positively sloped and exhibit increasing marginal rates of substitution, and assuming that consumption and leisure are normal, the Mirrlees-Spence single-crossing property applies:

$$-\frac{V^i_Y(C,Y)}{V^i_C(C,Y)} < -\frac{V^j_Y(C,Y)}{V^j_C(C,Y)} \qquad \forall\ w_i > w_j \tag{24}$$

for any bundle C, Y. Geometrically, at any consumption-income bundle, the slope of an indifference curve, $dC/dY|_{V^i}$, is less for a higher-ability person than for a lower-ability one. The choice of an optimal non-linear income tax structure is equivalent to the choice of consumption-income bundles for the various types of persons. The general problem of a planner whose objective function is welfaristic and satisfies the Pareto principle can be written:

$$\max \quad W\left(V^1(C_1, Y_1), V^2(C_2, Y_2), V^3(C_3, Y_3), \cdots V^n(C_n, Y_n)\right) \tag{25}$$

subject to

$$\sum_{i=1}^{n} N_i(Y_i - C_i) \geq R \tag{B}$$

$$V^j(C_j, Y_j) \geq V^j(C_i, Y_i) \qquad \forall\ i, j \tag{SS}$$

where N_i is the number of type i persons. The first constraint (B) is the government budget constraint, where R is some given revenue requirement. The second constraints (SS) are the self-selection constraints that must be satisfied among all pairs of households: each person must weakly prefer the bundle intended for them over any one else's bundle to preclude mimicking.

This problem can be used to trace out all Pareto-efficient combinations of household utilities by taking the objective function to be a weighted sum of utilities with weights being allowed to vary. But, for redistribution policy purposes, we may be interested in the set of Pareto-efficient allocations associated with an objective function that gives rise to some desire for redistribution from the better-off to the less well-off. For this purpose, the function $W(\cdot)$ can be taken to be quasi-concave in utilities.

Consider first the case in which there are two ability-types ($n = 2$). The single-crossing property implies that at most only one of the two self-selection constraints (SS) will be binding, and it is straightforward to show that with any quasi-concave social welfare function, the constraint applying to a type 2 person—$V^2(C_2, Y_2) \geq V^2(C_1, Y_1)$—will bind.[64] Moreover, $V^2(C_1, Y_1) > V^1(C_1, Y_1)$ since a given income corresponds to less labour supply for the higher wage household. Thus all allocations satisfying the self-selection constraint must satisfy $V^2(C_2, Y_2) > V^1(C_1, Y_1)$. Therefore, the second-best UPF, unlike the first-best one, will include only points such that type 2's are better off than type 1's.

Figure 5 illustrates some relevant features of the solution for the two ability-type case. For simplicity, we suppose there are identical numbers of the two types, and the government has no net revenue requirements ($R = 0$). Panel (a) depicts three allocations. The first one (L_1, L_2) is the laissez faire allocation where $C_i = Y_i$ for each household. Naturally, type 2's are better off than type 1's and (SS) is not binding in either direction. Imagine now the government implementing redistributive lump-sum (non-distorting) taxes contingent on incomes. This can occur until the allocation (S_1, S_2) is reached. At that allocation, the self-selection constraint on household 2 just becomes binding ($V^2(C_2, Y_2) = V^2(C_1, Y_1)$), and lump-sum redistributions can be carried no further. But further redistribution can be achieved beyond this point if distortionary taxes are allowed. The allocation (O_1, O_2) represents one such allocation, and one which satisfies the above planning problem for a planner with a quasi-concave

[64] As we saw earlier, the first-best outcome under a quasi-concave social welfare function will leave the low-ability person no worse off than the high-ability person. Since in the laissez faire the low-ability person is worse off, redistribution must go from the high-ability to the low-ability person. The self-selection constraint will become binding before equal utilities are achieved.

social welfare function. It can be implemented by an infinite number of tax schedules, or C, Y–schedules, a representative one of which is shown in panel (a). As can be seen, the marginal tax rate—one minus the slope of the C, Y–schedule—is discontinuous at the point O_1, and more generally in the multiple-type case at all allocations chosen by households other than the most able (except in the limit when the distribution of types in continuous). When we speak in what follows of the marginal tax rate, we simply mean the implicit marginal rate defined by the slope of the taxpayer's indifference curve at that point: more precisely, $T'(Y) = 1 + V_Y/V_C$.

FIGURE 5

Several features of this optimal income tax allocation are worth noting. The marginal tax rate applying to type 2's is *zero*, while that applying to type 1's is between zero and 100 percent.[65] Intuitively, the use of distortionary taxation allows the government to relax the self-selection constraint on type 2's in the following way. Begin at the lump-sum allocation (S_1, S_2) where the self-selection constraint just binds with lump-sum taxes. Since the opportunity cost of additional income is higher for type 1's than type 2's (reflecting the higher labour supply required to generate a given income), distorting income slightly downwards for type 1's by imposing a positive marginal tax, but reducing their after-tax income so their utility does not change makes the mimicking households worse off. Thus, the self-selection constraint is relaxed and redistribution can be pushed further. This is an important insight, for it shows that distortionary taxation need not be imposed as an exogenous constraint on the problem; it arises naturally as an optimal form of policy in a world of imperfect information. Moreover, as we shall see below, its role as a device for relaxing self-selection constraints induces us to look to other forms of distortionary policy which might serve the same purpose, an insight first noted by Nichols and Zeckhauser (1982).

In this pure-redistribution two-person case, the tax system is obviously progressive: high-ability types pay positive taxes, while low-ability types receive a subsidy. But, if the government has a revenue requirement sufficient to induce both persons to pay

[65]Formally, the Lagrangian for the government problem is

$$\mathcal{L} = W\left(V^1(C_1, Y_1), V^2(C_2, Y_2)\right) + \lambda[N_1(Y_1 - C_1) + N_2(Y_2 - C_2)] + \gamma[V^2(C_2, Y_2) - V^2(C_1, Y_1)]$$

The first-order conditions on C_2 and Y_2 yield $-V_Y^2/V_C^2 = 1$ (implying a zero marginal tax rate on type 2's). Those on C_1 and Y_1 yield

$$-\frac{V_Y^2}{V_C^1} = \frac{-\gamma \hat{V}_Y^2 + \lambda N_1}{\gamma \hat{V}_C^2 + \lambda N_1}$$

where a hat refers to a household of type 2 when mimicking a type 1. Since, by the property of diminishing marginal rate of substitution, $0 < -\hat{V}_Y^2/\hat{V}_C^1 < 1$, this entails that $0 < -V_Y^2/V_C^1 < 1$, implying a positive implicit marginal tax rate on type 1's.

taxes, there is no guarantee that the tax schedule is a progressive one. As in the first-best case, average tax rates may rise or fall with incomes.

Optimal tax allocations like (O_1, O_2) are obviously Pareto-inferior to at least some allocations on the first-best UPF. The informationally constrained, or second-best, UPF labelled tm is compared with the first-best UPF labelled PP for the two ability-type case in Figure 6. There is a range $s's$ around the laissez-faire allocation L which coincides with the first-best UPF. In this range, the self-selection constraints are not binding. In the range ts', redistribution goes from the low-ability types to the high-ability types, and the self selection constraint is binding on the low-ability types only. The point t is the 'maximax' point, the allocation that maximizes the welfare of the high-ability types. More relevant for our purposes is the range sm along which the self-selection constraint is binding for the high-ability types. For it is along this range that the solution must lie if a quasi-concave objective function is used. Since the self-selection constraint precludes any allocation which makes the low-ability types as well off as the high-ability types, the second-best UPF lies everywhere above the equal-utility locus, so neither the first-best utilitarian outcome U nor the first-best maximin outcome M can be achieved. The points u and m are the informationally constrained utilitarian and maximin allocations, and the range enclosed by them are all those that would be obtained from a quasi-concave objective function.

FIGURE 6

The analysis extends in a straightforward way to the multiple-ability case. Panel (b) in Figure 5 illustrates a typical case in which individuals are fully separated by ability-type. They line up by ability level, with self-selection constraints binding between each ability-type and the next lowest one (except for the lowest ability-type). Consumption, income and tax liabilities all rise with ability. A further complication arises in the multi-ability type and that is that partial pooling or bunching can occur whereby adjacent groups of ability-types choose the same C, Y–bundle (Guesnerie and Seade (1982)). In the bunching case, the self-selection constraints are also binding for adjacent ability-types, while consumption, income and tax liabilities are weakly increasing in abilities.[66]

[66]The circumstances under which bunching occurs has been studied for the continuous distribution case by Ebert (1992). In this case, each household faces the same continuous C, Y budget constraint. Incentive compatibility requires that each household maximizes its utility by choosing the point on the constraint intended for their ability type. The incentive compatibility constraints imposed on the problem must include both the first-order necessary conditions for the household's choice of C, Y bundle and the second-order sufficiency conditions. The latter turn out to require that incomes be non-decreasing. (See Myles (1995) for a careful exposition of this.) Bunching occurs when the second-order conditions are violated. Apparently, it is prone to occur at the bottom when the wage rate of the least able person is very low (Mirrlees (1971)) or when the welfare weights are relatively high on the lowest ability persons (Lollivier and Rochet (1983)), or at any point in the distribution when the

It may be optimal for one or more ability-types at the bottom not to be working, in which case they are also bunched, receiving the same C with $Y = 0$, so tax payments are negative.[67] Panel (c) of Figure 5 illustrates this case. It should be intuitively obvious that the lower the ability of the lowest income household, the more likely are there to be non-working households at the bottom. In the extreme, if the lowest ability types are completely non-productive ($w = 0$), as is often assumed (e.g., Mirrlees (1971)), there will necessarily be at least one ability-type not working at the bottom. And the lower the revenue requirements, the more likely is it that the non-negative income constraint will bind: reductions in the revenue requirement would cause the entire pattern of consumption-income allocations to shift to the left.

These results for discrete distributions have been presented for the case of a government with non-negative aversion to inequality, where redistribution goes from higher to lower ability persons. But, political economy considerations of the sort discussed in the previous section suggest that such schedules may not reflect those that would be chosen by decisive voters. As mentioned, Roell (1996) has supposed that voting might take place among the set of tax schedules most preferred by the self-interested members of the electorate, and one might expect higher income voters to prefer redistribution to go towards rather than away from themselves. A voter of ability type i will prefer the tax schedule that maximises $V^i(C_i, Y_i)$ subject to a revenue constraint and the relevant self-selection constraints; in other words, one which extracts as much revenue from the other voters as possible. Panel (d) in Figure 5 depicts the most preferred tax schedule for a middle-ability person in a five-ability world. For all persons above person 3, self-selection constraints are binding downward as above: redistribution goes from higher to lower ability persons. The top person has a zero marginal tax rate, while the next one faces a positive marginal tax. But, for persons of types 1 and 2, the self-selection constraints bind in the opposite direction since the poor redistribute to the less poor. The poorest person faces a zero marginal tax rate, while the tax for the second poorest is negative. As Roell's analysis indicates, the class of such most-preferred tax schedules may or may not satisfy the Smart-Gans single-crossing property.

These results tend to go through in the case in which there is a continuum of abilities, the case taken by Mirrlees (1971) and Tuomala (1990). There is still a zero marginal tax rate at the top, reflecting the absence of a self-selection constraint applying for the most able person. Marginal tax rates in the interior are all positive and less that 100 percent. In the absence of bunching at the bottom, the marginal tax rate for the lowest ability person is zero (Seade (1977)). If, however, there is bunching at the

density function of abilities is sharply decreasing (Ebert (1992)). Weymark (1986) has extended the analysis of the determinants of bunching to the discrete case.

[67] Formally, a non-negative income constraint must be imposed on all households. When it is binding on some households, these households will have zero incomes. Given that incomes are non-decreasing in abilities, these will be the lowest ability households.

bottom, the marginal tax rate facing the household at the end of the bunching interval is positive. Intuitively, since a positive marginal tax rate creates a deadweight loss the only purpose it can serve is to raise revenue from those higher up the distribution for redistribution to those below. But at the bottom of the distribution there is no-one further down to redistribute to, and so no equity gain to offset against the efficiency loss. If the poorest do not work, however, there is no such efficiency loss, leaving only the revenue-raising advantage of the positive marginal tax rate. As in the discrete case, utility is increasing with ability (reflecting the self-selection constraints), as are consumption, income and tax liabilities, except in the bunched intervals. But the analysis is very complicated and little in general can be said about the pattern of optimal tax rates, either marginal or average. That is because closed-form solutions for the optimal tax problem are generally not available except under special assumptions. For example, Diamond (1998) shows that if preferences are quasi-linear in comsumption, an explicit expression for the marginal tax rate can be derived in terms of the elasticity of labour supply, the distribution of ability types, and the form of the social welfare function. For this case, he finds that above some critical skill level, the pattern of marginal tax rates is U-shaped if the density of skills is single-peaked, the elasticity of labour supply is constant and skills follow the Pareto distribution above the modal skill level, with the minimum marginal tax rate occurring at the mode. But even here, average tax rates not specified.

Given the paucity of analytical results, simulation analyses have played a central role in developing a firmer sense of the likely nature of optimal non-linear tax schedules. Mirrlees (1971), in his classic study, calculated the optimal tax structure for the special case of a utilitarian social welfare function, Cobb-Douglas preferences in consumption and leisure, and a lognormal distribution of abilities. He found, famously, that the optimal schedule is then not very progressive: it approximates a linear tax with a relatively low marginal rate. Others have found that reasonable changes in specification can readily lead to greater progressivity. Stern (1976), simulating the optimal linear tax (characterised in equation (27) above) found the optimal marginal rate to be sensitive to the elasticity of substitution in consumption between consumption and leisure (which the Cobb-Douglas assumption restricts to unity): the easier the substitution—and hence the more elastic the labour supply—the lower the optimal marginal rate, which is as our earlier discussion would lead one to expect. Atkinson (1973) found that increasing the aversion to inequality of the social welfare function substantially changed Mirrlees' qualitative results (there being no aversion to inequality in utility in the utilitarian case considered by Mirrlees, of course): for example, with a maximin social welfare function the optimal tax is no longer roughly linear, and a much larger proportion of people at the bottom end of the ability distribution are optimally idle. This suggests that the progressivity of the optimal income tax depends critically on the inter-personal value judgements incorporated into the government's objective function.

Modifications of this classic Mirrlees-Stiglitz optimal income tax framework naturally modify the conclusions. One result which is of some importance for redistribution, and which is sensitive to the assumptions, is that the marginal tax rate for lower ability households should be positive. There are a number of model variations in which negative marginal tax rates, or wage subsidies, at the bottom are optimal, either to increase labour supply by low-skilled persons or to induce them into the labour force. These include: i) if the government's objective is non-welfaristic and relatively little weight is put on the leisure of the poor (e.g., if the aim is to reduce a poverty index involving consumption) (Oswald (1983); Kanbur, Keen and Tuomala (1994)); ii) if there are labour market distortions, such as efficiency wages or union power, which induce involuntary unemployment (Keen (1997)); iii) if skilled and unskilled labour are complementary inputs in the production process (Allen (1982)); iv) if there is a minimum amount of income that must be earned to enter the workforce (Boadway et al (1999)); v) if there are differences in productivity among workers in the market and non-market sectors (Beaudry and Blackorby (1997)); or, vi) simply if there are other tax-transfer programs which have an implicit positive tax rate on labour, such as a indirect tax on consumption.

4.3 Restricted instruments: Linear taxes

To implement the optimal non-linear income tax is a daunting task. Although the policymaker cannot identify households by ability-type, other informational requirements are highly demanding, including knowing household preferences and the distribution by types. Moreover, the setting is highly simplified relative to the real world: the economy is static, asset wealth is non-existent, households are homogeneous except for ability, there is no uncertainly, and so on. Although progress has been made to incorporate various of these elements into the optimal income tax formulation on a piecemeal basis, the kinds of tax structures obtained from the theory are typically much more complex than those used in the real world. policymakers simply do not fine-tune the tax structure to elicit truthful revelation the way the theory suggests they should.

Rate structures are typically relatively crude: many countries have piecewise linear tax systems with relatively few tax brackets of successively increasing marginal tax rates. These may be supplemented with indirect commodity tax systems with some differentiation of rates among major commodity groups (food, alcohol, luxuries, etc.). Why tax structures should be linear, piecewise or not, is not well understood. Simplicity itself may be a virtue, especially when administrative and compliance costs are taken into consideration, a point stressed by Slemrod (1990). Whatever the reason, the seeming inability of the policymaker to use other than relatively simple linear tax structures itself restricts the redistribution that can be achieved. We can summarise briefly some of the features of linear tax structures as they affect the redistributive

potential of the tax system.[68]

The most restrictive linear redistributive tax structure is one which relies entirely on differential commodity taxes. The early optimal tax literature characterised how the standard Ramsey tax rules would have to be revised to take account of distributive concerns (Diamond and Mirrlees (1971); Feldstein (1972)). Indeed the conflict was stark: roughly speaking, efficiency considerations suggested taxing goods in relatively inelastic demand, which tended to include those whose income elasticities of demand were relatively low, while distributional concerns suggested the opposite (Atkinson and Stiglitz (1972)). The summary statement was by Diamond (1975), who formulated a 'many-person Ramsey tax rule' which illustrated explicitly the equity-efficiency trade-off involved. In particular, per unit tax rates t_i on commodity X_i $(i = 1, \cdots, N)$ should satisfy the following conditions:

$$\frac{\sum_i t_i \sum_h S_{ik}^h}{H\bar{X}_k} = \sum_h \frac{b^h}{H}\frac{X_k^h}{\bar{X}_k} \quad k = 1, \cdots, N$$

where S_{ik}^h is the substitution effect between X_i^h and X_k^h for household h, H is the number of households, $\bar{X}_k$ is average household consumption of commodity k, and b^h is the net social marginal valuation of income going to household h, $b^h = W_h'\alpha^h/\lambda + \sum_i \partial X_i^h/\partial M^h$, where W_h' is the derivative of the social welfare function with respect to household h's utility, and λ is the Lagrange multiplier on the government revenue constraint. The left-hand side is analogous to Ramsey's proportional reduction (in compensated demands) rule, which in the single household case is the same for all commodities. Here, the proportionate reduction in compensated demands is smaller for commodities which tend to be more important for households with a high social valuation of income, reflecting the equity concern.

Apart from the part played by the many-person commodity tax rule in the development of the optimal tax literature, what possible interest might there be in it? Its main relevance would seem to be to the case of developing countries where, for administrative reasons, indirect taxes remain the mainstay of the revenue system. Although the indirect tax systems of such countries undoubtedly stray widely from many-person Ramsey optimality, the latter may nonetheless serve as a useful benchmark. Indeed, Ahmad and Stern (1984, 1991) have used the framework as a basis for determining welfare-improving tax reforms in India and Pakistan, starting from a tax system which is demonstrably different from an optimal one. Perhaps more to the point for our purposes, Sah (1983) has demonstrated just how impotent a system of indirect taxes is likely to be as a redistributive device. He shows analytically that, under the optimal indirect tax system, the proportional gain in real income (as measured by the

[68] Of course, although the rate structures themselves are relatively simple, the definition of the base and the structure of credits and exemptions can be relatively complex. These can be a source of differentiation among taxpayers which can enhance the extent of redistribution even for very simple rate structures. There has been surprisingly little analysis of this in the literature.

expenditure function at a representative set of prices) to the worst-off households is always less than the ratio of the maximum share of the budget among commodities for that household to the minimum average budget share among commodities across all households. That can be very small indeed: even for very large income differences among households, this ratio is of the order of 1.01–1.13 for the case of typical demand functions for the U.K. estimated using the Linear Expenditure System of demands.

The Sah analysis, like those of Diamond and Ahmad and Stern, is for a particularly restrictive form of linear taxation—indirect commodity taxes. Once households can be taxed directly, the redistributive potential of linear taxes can be enhanced. Consider the above many-person commodity tax setting. Suppose we enpand the policymakers instruments to include a poll tax or transfer on each household. As Atkinson and Stiglitz (1980) show, the policymaker can now combine the optimal indirect tax with a poll subsidy and obtain a general linear progressive tax system: a system of taxes on commodities combined with a lump-sum payment to all taxpayers. In the simplest case of one good and leisure, this is equivalent to a linear progressive income tax.[69]

The trade-off between efficiency and equity comes out particularly clearly for this case. Choosing the parameters of a linear income tax (exactly as in Section 3.1.2) so as to maximize a social welfare function $W[V^1, ..., V^H]$, for indirect utilities $V^i \equiv [(1-t)w^i, \alpha]$, subject to a revenue constraint:

$$\int tY_w(\alpha, t)dF(w) = \alpha \tag{26}$$

one finds from the first-order conditions that:

$$\frac{t}{1-t} = \frac{-cov[b, Y]}{\int Y\varepsilon_{LL}dF(w)} \tag{27}$$

where $b_w \equiv W'v_w/\lambda + tw\partial L/\partial\alpha$ is, as above, the net marginal social value of an additional unit of income and ε_{LL} is the compensated elasticity of supply of labour (Sheshinski (1972), Atkinson-Stiglitz (1980)). The numerator involves equity considerations, it being larger the more concavity there is in the social welfare function. The denominator is the efficiency term; it tends to make the optimal tax rate lower the more elastic is the compensated labour elasticity. Simulation results on the optimal linear income tax are reported by Stern (1976).

The assumption that the government can deploy only a linear income tax substantially restricts the utility possibilities open to the government. Moreover, it does not allow the government to exploit its redistribution objectives to the fullest, given the

[69] In a multi-commodity world, the optimal commodity tax structure will be uniform—so that only a linear income tax is needed—if leisure and all other goods are separable, with linear Engel curves from the former (Deaton and Stern (1986)).

information that is assumed to be available to it. To see this, consider Figure 7, which depicts linear progressive income tax options for the two ability-type case. The common budget constraint is hk, where the two ability-types choose allocations O_1 and O_2. Changing the budget constraint to the piecewise linear one $abcd$, where the segments ab and cd are both 45^o, allows household 2 to choose the preferred point e and household 1 to choose the preferred point b, while raising the same amount of tax revenue and not violating the self-selection constraint. This is a strict Pareto improvement. The source of the inefficiency of linear taxation is that all self-selection constraints are bound to be slack, indicating that an opportunity to improve the efficiency-equity tradeoff is not being exploited.

FIG 7

Variants of the simple linear progressive income tax can improve the efficiency of redistribution by expanding the UPF. Thus, Fair (1971) allows the income tax to be of a polynomial form. More recently, Sheshinski (1989) and Slemrod, Yitzhaki, Mayshar and Lundholm (1991) considered the extension to the two-bracket piecewise linear case. But, while these extensions relax the constraints on redistribution, the qualitative features of the linear case remain. In particular, the self-selection constraints remain slack, indicating a lost opportunity to exploit Pareto-improving tax structure changes.

We return then to the Mirrlees-Stiglitz optimal non-linear income tax, that which is optimal given the information available to the planner. But, as we have seen, because of the demands of incentive-compatibility, redistribution options are severely restricted relative to those available in the full-information first-best allocations. Can we do better than that? It turns out that once we leave the soothing world of the first-best, where lump-sum redistribution is the norm, and enter the world of the second-best where the policymaker is encumbered by imperfect information, not only is distortionary non-linear taxation 'efficient', so too may be a myriad of other non-standard policy instruments, and for somewhat similar reasons.

4.4 Self-selection and non-standard instruments

We have emphasised that nothing precludes the policymaker from implementing non-distorting redistributive taxes in a world with a discrete number of ability-types. But, increasing the marginal tax rate above zero serves to relax the self-selection constraints and thereby expands the equity-efficiency trade-offs available to the policymaker. This insight turns out to have much broader potential implications. If distortionary income taxation can relax the self-selection constraints, perhaps other distortionary policy instruments can as well, policy instruments that would not be used in a first-best world. There have been several examples of this in the literature. Some of them are as follows.

4.4.1 Public goods

So far we have assumed that revenue requirements are fixed, possibly at zero. Suppose now that in addition to redistributing from the better-off to the less well-off, the government requires revenue to finance a Samuelsonian public good, say G, where household preferences may now be written $V^i(C_i, Y_i, G)$. In a first-best world, it is well-known that the amount of G chosen would be that which satisfies the Samuelson conditions, $\sum_i V_G^i/V_C^i = p$, where p is the producer price of G in terms of C, which for simplicity we take as given.[70] Suppose though that the policymaker is imperfectly informed and must raise revenues using an optimal non-linear income tax. The problem for the redistribution-minded policymaker in the two ability-type case now becomes (Boadway and Keen (1993)):

$$\max_{C_i, Y_i, G} \quad W(V^1(C_1, Y_1, G), V^2(C_2, Y_2, G))$$

subject to

$$N_1(Y_1 - C_1) + N_2(Y_2 - C_2) \geq pG$$
$$V^2(C_2, Y_2, G) \geq V^2(C_1, Y_1, G))$$

The necessary conditions for this problem yield the standard optimal non-linear tax structure as discussed above, plus

$$N_1 \frac{V_G^1}{V_C^1} + N_2 \frac{V_G^2}{V_C^2} = p + \frac{\gamma \hat{v}_c^2}{\lambda} \left[\frac{\hat{V}_G^2}{\hat{V}_C^2} - \frac{V_G^1}{V_C^1} \right]$$

where, as before γ and λ are the Lagrange multipliers on the self-selection and revenue constraints, and 'hat' indicates a type-2 person mimicking a type 1. This condition reduces to the Samuelson condition, $\sum_i MRS_{GC}^i = p$, iff the mimicker's marginal evaluation of the public good relative to consumption equals that for the type 1 person being mimicked, $\widehat{MRS}_{GC}^2 = MRS_{GC}^1$. This will be the case if G and C are separable from leisure in the utility function (since a mimicker obtains the same G and C as the person being mimicked, but has more leisure time). If G is complementary with leisure, in the sense that its marginal valuation rises with leisure, $\sum_i MRS_{GC}^i > p$, so there is a tendency to under-provide the public good; and vice versa for the substitute case.

The intuition for this result is apparent and instructive. Suppose we start at an optimal non-linear tax allocation such that the Samuelson condition is satisfied, and assume that $\widehat{MRS}_{GC}^2 < MRS_{GC}^1$, so G and leisure are substitutes in the above sense. Increase

[70] We have chosen C to be the numeraire good here, but could have chosen leisure. In a first-best world that is of no consequence: a Samuelson rule will apply in either case. But, when distortions exist, the choice of numeraire affects the form of the optimality rule for public goods. For a discussion, see Boadway and Keen (1993).

G incrementally and adjust taxes so that $dT^1 = MRS^1_{GC}$and $dT^2 = MRS^2_{GC}$. This leaves V^1 and V^2 both unchanged, and keeps the government budget balanced. But mimickers are now worse off since they value the increment in G at less than MRS^1_{GC}, the additional tax levied. Thus, increasing G beyond the Samuelsonian optimal point relaxes the self-selection constraint and enables the policymaker to engage in further redistribution.[71]

4.4.2 Indirect taxation

The mix of direct and indirect taxes has long been an important policy issue. The informational approach to optimal redistributional policy uncovers a role for differential commodity taxes in relaxing the incentive constraints. Suppose that low-ability persons allocate their chosen income differently among goods than does a higher-ability mimicker. Then, using the same intuition as above, the self-selection constraint can be relaxed by imposing relatively high tax rates on goods that make up less of the low-ability household's consumption bundle, and relatively low ones (e.g., subsidies) on more preferred goods in such a way as to leave the household's welfare unchanged. This will make a high-ability mimicker worse off, relaxing the self-selection constraint and enabling more redistribution through the income tax to take place. Roughly speaking, as Edwards, Keen and Tuomala (1994) and Nava, Schroyen and Marchand (1996) show, this requires imposing relatively high tax rates on goods that are more complementary with leisure, a result reminiscent of the famous Corlett and Hague (1953) Theorem. If goods are weakly separable from leisure in the utility function, differential commodity taxation will be of no avail since households and their mimickers will differ only in leisure taken: they will allocate their identical incomes the same way across commodities (Atkinson and Stiglitz (1976)).

4.4.3 Quantity controls

If introducing price distortions can be welfare-improving by relaxing self-selection constraints, perhaps quantity controls can as well. In a seminal contribution, Guesnerie and Roberts (1984) showed that in a second-best economy with linear price distortions, imposing incremental quantity controls, either rationing or forced consumption, would generally be welfare-improving. In particular, if the distortions were commodity taxes, welfare would be improved by forcing a consumer to increase consumption of a good, assumed to be non-retradeable, whose consumption the tax system had discouraged, or to force a consumer to consume less of a good which the tax system had

[71] This result can be further interpreted as implying that the marginal cost of public funds is less than unity. In contrast, under linear income taxes, the marginal cost of public funds will typically exceed unity when taxes are linear (Browning (1976), Usher (1986)).

encouraged.[72] They subsequently suggested that this result could be used to condone the use of minimum wages, which are a form of labour market rationing (Guesnerie and Roberts (1987)).

The Guesnerie-Roberts analysis applied for linear tax distortions. As we have pointed out, linear tax regimes are inefficient in the sense that they do not exploit all the information available. The issue then arises whether quantity controls can still be welfare-improving when optimal non-linear taxes are in place. A series of papers have argued that they might be. It has been shown that in-kind transfers, such as education and health care, can enhance social welfare even when optimal non-linear taxation is in place (Blackorby and Donaldson (1988); Boadway and Marchand (1995); Blomquist and Christiansen (1995)). The in-kind transfers can be made anonymously available to all with the possibility of private supplementation, or the use of publicly provided goods may be mutually exclusive with private provision so that users of the former have to 'opt in'.[73] Another example concerns workfare, the requirement that transfer recipients perform some work. Given that the opportunity cost of working is less for low-ability persons than for potential mimickers, this might be expected to relax the self-selection constraint restricting redistribution to low-ability households. This turns out to be the case, at least as long as the productivity of workfare is not too low (Besley and Coate (1992); Brett (1998)). Finally, it has been shown that minimum wages, supplemented by unemployment insurance, can be welfare-improving as well even when the optimal non-linear income tax is in place (Marceau and Boadway (1994)). This literature on the use of public expenditures and pricing controls for redistributive purposes is still in its infancy. But, it is also somewhat suggestive and potentially relevant given the widespread use by governments of these type of policy instruments as well as other related ones, such as mandated private sector purchases (Summers (1989)).

4.5 Altering the information assumptions

The Mirrlees-Stiglitz model illustrates the implications for redistributive policy of the government not knowing certain features of the economy and of its inhabitants. It presumes that the government is ill-informed in one key respect—it knows only the aggregate distribution of household characteristics—but well-informed in another—it

[72] More precisely, let $d_n^h = -\sum_\ell t_\ell S_{n\ell}^h$ be the so-called index of discouragement of commodity n for individual h, where $S_{n\ell}^h$ is the compensated elasticity of demand for commodity n with respect to the price of good ℓ. If $d_n^h > 0$, a quantity control which forces individual h to increase their consumption of commodity n above that voluntarily chosen would be welfare-improving, and vice versa for $d_n^h > 0$.

[73] The conditions under which in-kind transfers are welfare-improving are similar to those under which preferential commodity taxation would be called for, leading to the question as to whether optimal commodity taxation alongside income taxation would eliminate the usefulness of in-kind transfers. But, Boadway et al (1998) have shown that, even when optimal commodity taxation is allowed for, the use of in-kind transfers can still be welfare-improving.

knows household preferences and can accurately observe incomes. Other informational assumptions are possible. If the government does not know even the aggregate distribution (or household preferences), there is an element of aggregate uncertainty to be resolved on top of the standard redistributive problem. But the nature of the limit to redistribution itself does not change.

On the other hand, the government may know more than just the aggregate distribution of abilities. For example, it may have access to a signal or 'tag' that is correlated imperfectly with underlying characteristics.[74] If so, the government can condition redistribution instruments according to the tag obtained by a person. Since tagging is imperfect, its use can involve both type I errors (failing to tag deserving persons) and type II errors (tagging undeserving persons). Segmenting the population into tagged and untagged groups allows the government to apply a separate redistributive tax-transfer system within each group, as well as redistribution from the untagged to the tagged group. The result is that more redistribution is made to the less well-off within the tagged group, but at the expense of less redistribution to the deserving who are mistakenly not tagged. If the tagging is costless, society is necessarily better off by its use since it expands the utility possibilities available. But if it is costly, the benefits of the additional information must be set against the costs. Moreover, as some observers have pointed out, the use of tagging may also raise questions of political feasibility (e.g., Sen (1995)). The more targeted are transfers towards the truly needy, the less support might be forthcoming from the excluded middle classes who constitute the bulk of voters. How telling this might be in the debate between those who would advocate targeting as a way of improving the efficiency of distribution and those who would opt for more universal schemes is surely a ripe item for future research.

The use of targeting also raises issues of the institutional delivery of redistributive programs. In principle, the tagging could be done through the tax system so that the program took the form of a (non-linear) negative income tax system, albeit one where the amount of the transfer was determined not only by income but also by other observable characteristics. But, in practice, most redistributive transfers are not delivered through the tax system, but through welfare agencies. Presumably, the reason is that the relevant tags are not readily observable to the tax collecting agencies, and the system of self-reporting that tax systems rely on is not suitable. Instead, the tagging is done by social workers who, through their own effort and the use of administrative resources, determine the category of eligibility into which applicants fall. In such a system, not only is the accuracy of the tag endogenous and dependent upon resources devoted to it, it is also dependent upon the effort and preferences of social workers. There is thus a conventional agency problem involved in

[74] The idea of improving the ability of redistribution programs to target people by the use of a tag was proposed by Akerlof (1978) and applied to disability programs by Diamond and Sheshinski (1995) and Parsons (1996). For an extended discussion of it in the context of redistributive programs in developing countries, see van de Walle and Nead (1995).

the tagging process, whose costs must be factored into determining not only whether tagging should be done, but also how many categories of targeting should be used (Boadway et al (1999)).

4.6 Cheats and liars

The standard optimal redistributive tax theory just discussed assumes that the otherwise ill-informed government can observe all taxpayers' incomes. But, income tax liabilities are typically based on self-assessment by taxpayers, so the accuracy of this procedure relies on the taxpayers truthfully reporting their incomes. The extent to which truthful reporting occurs depends not only on the standards of behaviour of the community,[75] but also on the monitoring and sanctioning activities of the tax authorities. In theory, one might suppose, following Becker (1968), that penalties for evading taxes could be set high enough to eliminate it at minimal cost to the administration. In practice, such maximal sanctions are rarely observed for tax evasion, let alone for other criminal acts. Various arguments have been put forward for the absence of maximal sanctions: there may be errors in conviction; sanctions may be costly to impose and the costs may increase with the level of sanction; criminals may be able to engage in costly avoidance activities which reduce the probability of being caught; there may be imperfect information about the probability of apprehension or about whether acts are subject to sanctions; if law enforcement is general so that all crimes are deterred with the same probability, and if there is a limit to the maximal sanction so that the general probability of apprehension is not too small, it may be optimal to use less than maximal sanctions for lower gain crimes to prevent over-deterrence of these crimes; if criminals are risk averse, and if some crimes are 'good' ones in the sense that the private gain to the criminal exceeds the social cost, it might not be optimal to impose a maximal sanction with a small probability of detection; and, there may be purely ethical arguments against the state imposing maximal sanctions. Whatever the reason, it is typically simply assumed in most of the tax evasion literature that there is some limit to the penalty for tax evasion, a limit that is below the maximal sanction.[76]

The presence of tax evasion that this implies is a further limit to redistribution. At the simplest level, for a given probability of detection and penalty, the incentive to

[75] Gordon (1989) and Bordignon (1993) study how social norms affect taxpayer honesty in reporting their incomes.

[76] The issue is discussed in more detail in Cowell (1990). It is also possible that admitting some tax evasion may be useful for other reasons, as Stiglitz (1987) argues. For one thing, inducing some tax evasion by low-ability households could relax the self-selection constraint if it makes it more costly (in expected terms) for higher-ability ones to mimic them. But, this requires low-ability persons to be less risk-averse, which may not be plausible. Alternatively, allowing some tax evasion is analogous to imposing random taxation. This might be efficient if the government aggregate revenue function is not concave in the tax rate, something which is a distinct possibility in second-best problems.

evade increases with the marginal tax rate: the expected benefit from reporting one less pound of income is the tax revenue saved if not caught. This suggests that the revenue responsiveness of an increase in the tax rate will be reduced by the presence of evasion, so that the MCPF is increased (Usher (1986)). This would seem to increase the efficiency costs of redistribution, and thereby further worsen the efficiency-equity trade-off. More formal treatments of optimal redistributive policy in the presence of tax evasion exist in the literature. Marhuenda and Ortuño-Ortín (1997), building on the model of Chandar and Wilde (1998), investigate the form of tax and auditing functions which induce truthful self-reporting of incomes to the tax authorities when incomes are exogenous, penalty function satisfy certain social norms, and the government objective is to redistribute income. In particular, it is assumed that 1) penalties cannot exceed agents' true incomes; 2) audited agents who are dishonest pay at least the same amount in taxes and penalties as honest agents of the same income level; and 3) the 'punishment fits the crime' so the penalty function is continuous and related to the severity of the crime (rather than being a maximal sanction in the sense of Becker (1968)). They find, surprisingly, that even when incomes are fixed, average tax rates are decreasing with income, that is, the tax system is not progressive.[77]

Cremer and Gahvari (1996) investigate the more general case of the optimal non-linear income tax along with penalty structures and audit strategy when incomes are endogenous. The penalty is constrained not to exceed a taxpayer's income. The government must induce the truthful reporting of income as well as satisfying the standard self-selection constraints. The form of the income tax structure for the two ability-type case they investigate is similar to the standard case where incomes can be observed: the marginal tax rate at the top is zero, while that for the low-ability types is zero as long as the self-selection constraint is binding. But, the self-selection constraint may not be binding: redistribution may be limited not by the fact that high-income persons can mimic the true income of the low-ability persons, but by the fact that high-wage persons must be precluded from reporting their incomes to be that of the low-ability persons even though they earn more. If so, the marginal tax rate is zero for both persons. In any case, there is no auditing at the top, since with a zero marginal tax rate there is no saving in tax payments from misreporting incomes. Only low-ability persons are audited. Moreover, honest reporting should be rewarded, so those audited and found to be honest should pay lower taxes than those not audited.

[77] One problem with the analyses of both Marhuenda and Ortuño-Ortín and Chandar and Wilde is that they assume tax liabilities must be non-negative. This would seem to rule out an equal lump-sum payment to all taxpayers, which could turn a tax system from regressive to progressive even if marginal tax rates were decreasing. Cremer, Marchand and Pestieau (1990) analysed the optimal linear income tax in a similar setting in which the penalty was fixed, but audit probabilities could vary with reported incomes. The optimal audit and linear tax system which induces truth-telling is one in which audit probabilities increase with reported incomes up to a cut-off level. Beyond that, there is no audit. The tax system has piecewise linear marginal tax rate: up to the audit cut-off, the marginal tax rate is positive, but beyond that it is zero. Thus, at least over the lower part of the income range, the tax will be progressive.

Unfortunately, as in other optimal income tax problems, little can be said about the progressivity of the tax. But, given that the inability to observe true income further constrains government, presumably the progressivity of the tax is lower than in the standard optimal income tax case.

The under-reporting of income is not the only form of illegal behaviour associated with the tax system. In some settings, tax administrators themselves may engage in corrupt practices. Tax collectors may collude with taxpayers to under-report income, with the gain being shared by the two parties (Besley and McLaren (1993), Flatters and MacLeod (1995)). Or, if their own income depends in part on the revenue they collect, tax collectors may credibly engage in extortion by demanding payments from taxpayers under threat of over-reporting their income. Hindriks, Keen and Muthoo (1998) show that, for a quite realistic class of tax, penalty and incentive schedules, the distributional impact of such practices is clear-cut: the effective tax system—taking account not only of tax payments but also bribes and penalties—is unambiguously less progressive, in the sense of implying a less equal distribution of net income, than the statutory tax system. Moreover, since the potential gain from understating income by one pound increases with the level of income under a progressive tax system, in order to implement such a scheme honestly it may be necessary to offer inspectors an incentive payment that increases with the tax they collect, a cost of progressive taxation quite distinct from those usually emphasized.

4.7 Time consistency

Our discussion so far has concentrated mostly on various forms of imperfect information as the reasons why the economy's second-best UPF shrinks relative to the first-best one. But, once one takes the principal-agent view of government's relationship to its citizens seriously and allows for a sequencing of decisions, another important source of restriction on the UPF arises. In an intertemporal world, if the government is restricted to using distortionary tax instruments, second-best policies are time-inconsistent: even the second best UPF is not attainable.[78] The intuition for this is straightforward. Suppose households decide both how much wealth to accumulate at the beginning of a sequence of periods and how much labour to supply in each of the periods. The government might calculate its sequence of optimal tax rates on, say, labour and capital incomes at the beginning of the first period, taking account of the responsiveness of household savings and labour supply decisions. If the government announced these tax rates, households acted on them and the government stuck by its announcement, the second-best optimal tax allocation would result. But, once households have taken their savings decisions, if the government could re-optimize, it would recognize that wealth so accumulated was now fixed. It would prefer to renege

[78]This was demonstrated by Hillier and Malcomson (1984) and Calvo and Obstfeld (1988). An early application to tax policy was by Fischer (1980), followed by Rogers (1987).

on its announced second-best policies, increasing taxes on capital income and reducing taxes on labour income, which is still variable. The second-best policies would not be time-consistent (sub-game perfect). Notice that this is not a consequence of irrationality on the government's part: a fully rational government could not prevent itself from reneging on announced second-best optimal tax rates.

In contemplating equilibrium responses to the time-inconsistency of second-best policies, it is natural to focus on policies that are time-consistent: policies, that is, which the government will not find it *ex post* optimal to renege on. In a simple economy with homogeneous consumers, optimal time-consistent policy equilibria will be clearly welfare-inferior to second-best ones. They will involve relatively high rates of tax on capital income, rates that are rationally anticipated by households and lead to lower wealth accumulation than in the second-best optimum (Fischer (1980)). In an economy with heterogeneous agents in which the government has a redistributive motive, time-consistent tax policies are also likely to restrict the equity-efficiency trade-off significantly. Such tax policies will tax wealth owners 'too much', and that is likely to lead to excessive redistribution given the government's goal: if past accumulated wealth is highly unequally distributed, an egalitarian government cannot prevent itself from imposing highly redistributive taxes on it. But, the end result of this may well be welfare-reducing because the redistributive tax policy, being anticipated by households, will induce a reduction in wealth accumulation. Boadway et al (1996) construct a model in which wealth accumulation takes the form of human capital investment, with different individuals able to obtain differing returns from a given amount of such investment. Because the government cannot avoid taxing the returns to human capital investment in a highly redistributive way, the time-consistent outcome can actually be Pareto-inferior to the *laissez faire* allocation.

As in the case where the limits to redistribution result from binding self-selection constraints, the problem of time inconsistency may induce the government to search for otherwise unconventional policy instruments. We have seen (in Section 2.2.2) how, in the context of the samaritan's dilemma, it can rationalize mandatory insurance and pension schemes. Clearly too there are potential implications for taxes on saving and investment. Given that time inconsistency results in too little wealth accumulation, governments may resort to up-front incentives for saving or investment, or even quantity controls. Indeed, a cursory glance at real world policies reveals not only that capital tax rates are much higher than optimal tax reasoning might suggest, but also that they tend to be accompanied by policies that mitigate their effects, such as investment tax credits and subsidies, incentives for human capital investment and retirement savings (including mandating these types of wealth accumulation) (Kotlikoff (1987)), tax holidays (Vigneault (1996); Wen (1997)), and lax enforcement of capital income tax evasion (Boadway and Keen (1998)).

4.8 Interjurisdictional Mobility

Mobility of households across jurisdictional borders poses a constraint on redistribution in a decentralized federation or an economic union (a recent survey being provided by Cremer et al (1996)). The basic argument was put by Boskin (1973) and Pauly (1973). Suppose there are several jurisdictions, say local governments, each populated by some 'rich' and some 'poor' persons. If the rich are altruistic towards the poor, they will agree collectively through their local government to make transfers to the local poor financed by taxes on themselves. If persons were not mobile across localities, the efficient amount of redistributive transfers would be made in each jurisdiction (and would differ according to local preferences). Now suppose the poor are mobile across jurisdictions: they will respond to different transfer levels offered by different local governments. If local jurisdictions take into account the effect of their transfer policies on the number of poor in their own jurisdiction, they will have an incentive to reduce the level of transfer payments to the poor. The movement of one poor person to another jurisdiction will result in a saving of the transfer that would be made to the poor person, but, since the poor person will obtain a transfer in the destination jurisdiction, there will be no corresponding reduction in altruistic benefit. Thus, there will be a tendency to free ride on neighbouring jurisdictions, and the Nash equilibrium will result in too low a level of transfers from an efficiency point of view. The lower the cost of moving, the more responsive will migrants be to transfer differences among jurisdictions, and the greater will be the severity of the free-riding problem. The argument applies whether or not altruism applies only to the poor located within one's jurisdiction. It also applies more generally for local government objective functions other than those based on altruism.

It may be, however, that while labour mobility may make redistribution harder it may also make it less necessary: it enables individuals to self-insure by moving elsewhere in the face of unanticipated shocks (Wildasin (1998)). Moreover there do exist models in which a reduction in the extent of redistribution brought about by labour mobility is actually socially desirable.[79] The general presumption, however, has been that free-riding is likely to lead to less redistribution than is desirable; and much attention then focussed on the question as to how to redress this bias.

The most obvious is to assign redistribution to some level of government that overarches the range of mobile labour. This is a classic prescription for the allocation of

[79] Janeba and Raff (1997) show that this can be the case, for example, in a voting model with more distinct jurisdictions than ability types. Each jurisdiction, suppose, can levy a linear income tax on its residents. Preferences are assumed to be such that—amongst the class of linear tax schedules—expected utility (behind the veil of ignorance) is maximized by zero taxation. Such an outcome is realised if labour can move, since in equilibrium there is only one ability type within each jurisdiction (and hence no redistribution). But if labour cannot move—and ability types are not fully sorted by region in the intial situation—then a standard median voter story means that some redistibution will occur.

tax powers in a federation (Musgrave (1959); Oates (1972)).

An alternative approach builds on the observation that the tendency to compete away transfers to the poor is a form of inter-jurisdictional externality that can be corrected by a matching grant from higher to lower levels of government. Wildasin (1991) constructs a model of a central government and several lower-level jurisdictions, each of which are populated by some immobile taxpaying households and some mobile poor households. The poor supply a fixed amount of labour and obtain market income from the value of their marginal product. The taxpayers own fixed factors in their jurisdiction of residence and obtain market incomes equal to their rent, which in turn depends upon the number of poor workers. The taxpayers care about their own consumption and the per capita consumption of the poor in their jurisdiction. Each lower-level jurisdiction acts in the interests of the local taxpayers and makes transfers to the poor to satisfy their altruistic objectives. Since the poor are mobile, there will be an interjurisdictional externality associated with local redistribution: an increment in transfers by one jurisdiction will induce an inflow of poor from other jurisdictions, reducing the wage and consumption of the local poor and increasing them in other jurisdictions. In a Nash equilibrium, each jurisdiction will ignore the benefit its transfers yield for taxpayers in other jurisdictions, and so will under-redistribute. Wildasin shows that when the poor are perfectly mobile, optimal central government policy is to subsidize local jurisdictions' transfers by matching grants whose rates are such that in equilibrium *all* jurisdictions make the same transfers to each poor person (despite the fact that they may have different degrees of altruism). This requires higher subsidy rates for jurisdictions with lower transfers, and vice versa. In this way the redistribution function can be efficiently decentralized to the lower-level jurisdictions.

Boadway, Marchand and Vigneault (1998) obtain a related result in an explicitly federal context. Both both central and lower-level governments engage in redistributive activities, using a linear progressive tax on earned income. The two levels of government are assumed to share the same objective function. It emerges when there are two income-classes of persons, redistribution can be fully decentralized to the states provided the federal government makes appropriately differential per capita lump-sum transfers to the two states.

These approaches to the problem all presume, however, some over-arching federal authority. Matters are evidently more difficult in circumstances such as that of the European Union, with no central authority mandated to act in such matters. One approach which has been suggested in such a context is that taxation (and entitlement to benefit) be determined not by residence but by nationality (Sinn (1989)). Only the US and the Philippines have sought to tax their citizens wherever in the world they happen to reside; doing so obviously raises considerable practical difficulties (the information-sharing requirements are likely to be formidable), and wider difficulties too concerning the right to renounce one's nationality (Pomp (1989)). Though not entirely far-fetched—arrangements are already in place to pay pensions to nationals

who retire in other member states, for instance—this seems a distinct prospect.

It is not only the mobility of labour that jeopardizes redistribution. The more dramatic developments in recent years has been the increased mobility of capital: and with residence-based taxes proving increasingly hard to enforce (being weakened by evasion and avoidance which some jurisdictions find it in their interests to promote) and source-based taxes looking vulnerable to downward tax competition, the pressures towards lower taxation of capital is pronounced. The distributional implications of this are not entirely clear—recipients of capital income include pensioners with low lifetime incomes, for instance—but the effects may clearly be substantial.[80]

4.9 Altruism and intergenerational redistribution

As noted earlier, while altruism can provide some rationale for redistribution it also constrains the amount of redistribution that can be achieved: see section 2.2.1.

5 Measuring Redistribution

The previous two sections have emphasised some of the factors tending to compromise the extent to which the objectives of redistribution can be achieved in practice. Not only are there significant limits to the utility possibilities, or the possibilities for achieving other non-welfaristic objectives, that can be achieved because of informational and other restrictions, but also the political process itself may conspire to confer high weight on the amount of redistribution preferred by selected groups in society. How these tensions resolve themselves is an empirical matter, and there is a disparate literature that attempts to measure the amount of redistribution that occurs in the real world. We cannot do full justice to that literature, given the many methodologies that can be used and the many countries and time periods to which the methodologies can apply. Our purpose is limited to outlining the main procedures that have been applied.[81]

[80] There is a large literature (surveyed in Keen (1996)) on the implications of, and possible responses to, current pressures on capital income taxation. For a dissenting view on the likelihood of capital taxes vanishing, Mintz (1994).

[81] Attention is confined too to the evaluation of redistribution by the public sector so that we do not consider, for example, the extent and effect of charitable gifts, bequests and the like. For a discussion of this aspect of redistribution, see Rose-Ackerman (1996).

5.1 Methodological issues in assessing redistribution

By its very nature, the measurement of the amount of redistribution implicit in existing or alternative fiscal policies is a counterfactual exercise. It involves measuring how a given set of policies affects different classes of households compared with those of some benchmark policy that is in some sense distributionally neutral. Quite what that might mean is far from obvious. One approach is to take as counterfactual a situation in which there is simply no government. There is an evident difficulty in applying this concept of *absolute incidence*. For the fiscal system includes not only the tax-transfer system, whose redistributive effect we take it one is most interested in measuring, but also public expenditure on goods and services. The latter will also affect real incomes to some degree, but since the goods are typically unpriced and have varying degrees of publicness associated with them, their redistributive effect is hard to measure. (Indeed if it were easy to measure the value consumers place on them, public provision would hardly be warranted).

An alternative approach—that of *differential incidence*—is to conceive of holding constant the observed levels of public expenditure but instead financing that provision in a distributionally neutral way. The difficulty of defining a distributionally neutral tax-transfer system replaces the difficulty of measuring the distributive effects of public expenditures on goods and services. Some apparently see appeal in taking the benchmark system to be an equal per capita tax—a poll tax—levied on all households; just as all typically pay the same price for each unit of a private good, so perhaps we should see it as natural that all pay the same price for public services. The failure of the poll tax in the UK, however, suggests that many see taxes of this kind as far from equitable. A much more common notion of distributional neutrality is that of a proportional tax system: one, that is, which takes the same share of income from all households. Such a system is the only one with the feature that the Lorenz curves of before-tax income, after-tax income and tax payments all coincide.

5.1.1 Progressivity comparisons

It is the notion of a proportional tax as being distributionally neutral, combined with the appeal of the Lorenz criterion, that underlies the standard definition of a progressive tax as one characterized by an average tax rate that everywhere increases with pre-tax income, since that ensures—as a corollary of a result expanded on below—that whatever the distribution of pre-tax income, post-tax income will be distributed more equally (in the Lorenz sense); and tax payments will be distributed less equally.

One wants of course to go further than statements about the existence of redistribution or—the aspect we focus on here—of progressivity in the tax schedule.[82] One would

[82] An excellent account of the concept and measurement of progressivity is provided by Lambert

like to make statements to the effect that redistribution is greater under one system than another, or that one tax is more progressive than another. The natural way to approach this is by looking for Lorenz dominance across the relevant distributions. Opinion differs, however, as to whether it is the distribution of tax payments or of after-tax incomes that should matter for progressivity comparisons. How progressive. for example, is capital gains tax? Payments are generally heavily concentrated amongst high income groups; but they are also sufficiently low that the tax has relatively little effect on the distribution of after-tax incomes. These two views of progressivity can give different answers: it may be, for example, that a reform generates a more equal distribution of net income but the Lorenz curves for net income cross. This though is a rather sterile dispute. The more fundamental questions are how to make comparisons when Lorenz curves cross and whether there are conditions under which one can be sure they will not. On the former, much attention has been given to the development of global measures of progressivity that will enable a complete ordering of schedules.

An answer to the latter question is provided in a result variously derived by Fellman (1976), Jakobsson (1976) and Kakwani (1977). This shows that, whatever the distribution of underlying income Y, the Lorenz curve of a schedule $a(Y)$ Lorenz dominates that of $b(Y)$—assuming both schedules to be everywhere strictly positive—if and only if the elasticity of a everywhere exceeds that of b. Thus one tax is more progressive than another in the sense of inducing a less equal distribution of tax payments iff it has a greater 'liability elasticity' (elasticity, that is of tax payments with respect to pre-tax income) at all income levels. Similarly it is more progressive in the sense of generating a Lorenz-dominating distribution of net income iff it has a lower 'residual income elasticity.' Comparing the distribution of tax payments and net income under a schedule $T(Y)$ with that of a proportional tax $t \cdot Y$, for instance, one finds the former to be unambiguously more progressive if the marginal tax rate exceeds the average, so that the average rate is everywhere increasing: exactly the definition of progressive tax given earlier.

In many contexts of course these elasticity conditions will not hold; the distributional impact of taxation will then depend on the underlying pre-tax distribution. Moreover, the application to the distribution of tax payments requires an unappealing assumption that the tax payments are always strictly positive. This is clearly not the case; as a consequence, that result cannot be applied to such basic questions as whether an increase in tax allowances leads to an increase in progressivity: for a key aspect of such a reform is that it takes some people out of tax altogether (an apparently progressive move), whilst at the same time conveying an absolute benefit that is greatest to those facing the higher marginal tax rate (which is apparently regressive). Keen, Papapanagos and Shorrocks (1998) extend the Feldman-Jakobsson-Kakwani result to deal with such situations. The effect of increasing tax allowances is seen to depend on a condition not previously encountered in this area: for such an increase to increase

(1993).

progressivity (in terms of the distribution of tax payments), it is sufficient[83] that the log of the tax schedule is everywhere log concave: for then the proportionate reduction in tax liabilities implied by an increase in the allowance falls with the level of pre-tax income.

Unambiguous progressivity comparisons cannot be made if these conditions fail. There may nevertheless be instances in which distributional inferences can be made from the shapes of tax schedules conditional on the distribution of pre-tax income. Hemming and Keen (1983) show, for example, that if two schedules raise the same revenue and cross only once then the net income distributions they induce can be Lorenz-ranked. Extensions of this result to multiple crossings are discussed in Lambert (1993).

All this assumes, however, that there are no behavioural responses to taxation. If there are, then—as Allingham (1979) shows by example—one schedule may be more progressive than another in the Feldman-Jakobsson-Kakwani sense and yet lead to a less equal distribution of net income.

5.1.2 Incidence and deadweight loss

Behavioural responses also mean that the impact of taxation on a household's real income cannot be inferred from the amount of tax that it actually pays.[84] Consider for instance the evaluation of the differential incidence of a system composed entirely of indirect taxes. Actual tax paid in the observed economy is $(\boldsymbol{Q}-\boldsymbol{P}).'\boldsymbol{X}(\boldsymbol{Q},U)$, where $\boldsymbol{Q}$ and $\boldsymbol{P}$ are the vectors of consumer and producer prices respectively (so that $\boldsymbol{T}=\boldsymbol{Q}-\boldsymbol{P}$ is the vector of commodity taxes in specific form), and $\boldsymbol{X}(\boldsymbol{Q},U)$ is the vector of Hicksian demand functions; U is the level of utility attained in the taxed equilibrium, so that, characterizing preferences by an expenditure function $E(\boldsymbol{Q},U)$ and indirect utility function $v(\boldsymbol{Q},Y)$, we have $U=v(\boldsymbol{Q},E(\boldsymbol{Q},U))$ Denote by T^* the taxes paid in the counterfactual situation in which the same revenue is raised as in the observed situation but by a proportional tax on income; and suppose too that this means a tax on endowment income, so that T^* is a lump sum tax. Assuming pre-tax endowment income to remain unchanged, so that $E[\boldsymbol{P},v(\boldsymbol{P},Y-T^*)]=Y-T^*$, it is readily seen that

$$E[\boldsymbol{P},v(\boldsymbol{P},Y-T^*)]-E[\boldsymbol{P},v(\boldsymbol{P})]=\{(\boldsymbol{Q}-\boldsymbol{P})'\cdot X(\boldsymbol{Q},U)-T^*\}+L(\boldsymbol{Q},U) \quad (28)$$

where $L(\boldsymbol{Q},U)\equiv E(\boldsymbol{Q},U)-E(\boldsymbol{P},U)-(\boldsymbol{Q}-\boldsymbol{P})'\cdot X(\boldsymbol{Q},U)$. The real income loss that an individual experiences from paying the observed indirect taxes rather than

[83] And necessary too if the result is to hold for all initial positions.

[84] By 'pays' we refer not to legal liability—the question of formal incidence—but the amount by which the value of an individual's consumption bundle at consumer prices exceeds the value at producer prices.

trading at current producer prices and instead paying the benchmark proportional tax is thus the sum of two effects. The first is simply the change in the amount of tax paid. The second, L, is a measure of the deadweight loss from the tax system:[85] it indicates the excess of the amount that the consumer would be willing to pay in order to trade at producer rather than consumer prices over the amount of tax they would thereby save. It is readily seen that $L \geq 0$, the equality holding only if there are no substitution effects in consumption. More generally, calculating comparative tax payments under actual and benchmark situation will understate the real income loss that an individual suffers from taxation. Some households, for example, who pay more tax under the proportional tax system may actually enjoy such a reduction in deadweight loss that they are actually better off than under the initial system. How this bias will affect of overall picture of the impact on the distribution of real income depends on how deadweight loss varies with the level of the endowment, about which there are few general results. It is clear, nevertheless, that focussing simply on tax payments may give a misleading impression.

But the pattern of tax payments—defined as above, rather literally, as the difference between the value of a household's observed net trades $\boldsymbol{X}(\boldsymbol{Q}, U)$ at consumer and producer prices—may nevertheless be a matter of considerable interest in its own right. For in measuring the excess of the value of the resources they provide over the resource cost of servicing their demand, it indicates precisely an individual's net contribution to the economy. As argued by Hicks (1946) and pursued by Dilnot, Kay and Keen (1990), this seems to be a quantity that people care about; the latter give the example of the lack of sympathy likely to be extended to a multi-millionaire who notes that although he paid no tax this year, it had been a considerable inconvenience to arrange his affairs so as to bring about this outcome. It is a quantity, moreover, whose calculation requires absolutely no assumption about the shifting issue to which we turn shortly: it depends only on prices and quantities in the initial situation, not those of any counterfactual equilibrium. Indeed we shall see that common incidence assumptions are tantamount to simply measuring tax payments thus defined; and in this sense a concern with the distribution of tax payments as such may provide a more coherent rationale for many incidence studies than does the concept of differential incidence itself.

A key assumption underlying (28)—though not of course the argument for an intrinsic interest in tax payments as such—is that producer prices are the same when the consumer trades at those prices as when he trades at consumer prices. Though perfectly coherent in itself, there is an evident problem in maintaining this assumption when assessing the redistributive impact over the entire set of consumers: if in practice distorting taxes were removed, producer prices would generally change from their level in the taxed equilibrium, to $\boldsymbol{P}^*$ say. Proceeding as before, and allowing too for the possibility that the value of the endowment may change (from Y to Y^*)—perhaps be-

[85] The desirable properties of this particular measure are shown by Kay (1980).

cause the individual sells part of the endowment at producer process, perhaps through asset price effects— one now finds that the change in real income in moving to the counterfactual state is

$$E[\boldsymbol{P}^*, v(\boldsymbol{P}^*, Y^* - T^*)] = \{(\boldsymbol{Q} - \boldsymbol{P})' \cdot \boldsymbol{X}(\boldsymbol{Q}, U) - T^*\} + \{Y^* - Y\} + L(\boldsymbol{Q}, U) \quad (29)$$

Detailed incidence studies of the kind discussed below often focus on the amount by which the cost of an individual's observed consumption attributable to facing the prices of the taxed state rather than counterfactual equilibrium exceeds the tax they would pay under the counterfactual system. This is precisely the amount $(\boldsymbol{Q} - \boldsymbol{P})' \cdot \boldsymbol{X}(\boldsymbol{Q}, U) - T^*$ in (29): which is seen to understate the welfare loss suffered by the household by the amount of the deadweight loss plus the reduction in the value of the endowment induced by the tax system.

In practice, excess burdens are accounted for to varying degrees in incidence studies. Those which are based on fully specified general equilibrium models of the sort discussed below automatically take account of excess burdens in calculating real income effects. Of course, the magnitude of these excess burdens will be model-specific. Those which rely more on assigning tax and transfer burdens to income classes according to some assumed rules of shifting typically ignore excess burdens: they assign all taxes and transfers to income classes, no more.

In any event, implementation, whether with a view to a full welfare analyses or for calculation only of pseudo tax payments, requires estimation of the prices $\boldsymbol{P}^*$ and endowment incomes Y^* of the counterfactual equilibrium, which brings us into the area of incidence analysis. This is a vast topic, much of it excellently surveyed by Atkinson (1994), Atkinson and Stiglitz (1980), Kotlikoff and Summers (1987), and Jha (1998). Here we focus only on recent developments in the analysis of incidence under various forms of imperfection, which may prove to have significant implications for assessing the distributional impact of the tax-benefit system.

5.1.3 Imperfect markets, unemployment and tax incidence

Recent years have seen especially striking progress in two broad areas of incidence analysis. One is the analysis of intertemporal issues; we review work on this in Section 5.3. The other, on which we focus, is the analysis of shifting issues in imperfect markets.

Imperfect product markets

Commodity taxes can have surprising effects once one leaves a world or perfect competition: they may be 'over-shifted,' with consumer prices increasing by more than

the amount of the tax; and they may even increase profits. The intuition behind these results—established by Seade (1985) and Stern (1987), but anticipated by Cournot (1838) and analysed by Musgrave (1959)—is straightforward. The possibility of over-shifting follows quickly from the the monopolist's first order condition: $(Q - T - MC)/Q = 1/\epsilon(Q)$, where $\epsilon(Q)$ is the elasticity of demand and $\epsilon < 1$ for a maximum. Assuming marginal cost MC to be constant, a tax increase dT is marked up by the factor $1 - (1/\epsilon)$ to produce an even greater increase in Q. The possibility that a tax increase may even lead to higher net profits arises from its effect in bringing about a coordinated contraction in the output of all firms, a contraction that their isolated self-interest fails to produce. That is, it enables firms to deal with the externality that each imposes on the other when it expands its output; profits increase if this gain through mitigation of external effects (which cannot arise in the monopoly case) dominates the direct effect of a higher tax rate.

Product quality also becomes an issue once the assumption of perfect competition is relaxed. Differential indirect tax structures may then have implications of product quality that in turn have distributional consequences. Predominantly ad valorem taxation, for example, is likely to lead to lower product quality than is predominantly specific: under specific taxation, the producer need only raise the consumer price by \$1 to recoup the cost of a \$1 quality improvement; under a 50 percent ad valorem tax, on the other hand, it must raise the consumer price by \$2. Myles (1988) establishes, moreover, the surprising result that a utilitarian may wish to subsidise a high quality good consumed by the better off; the intuition being the t since the rich choose to purchase the higher quality good but the poor do not there must be a point at which the marginal utility of income of the rich exceeds that of the poor.[86]

The empirical work to which these theoretical developments lead has not found strong evidence of over-shifting (Besley and Rosen (1994), Baker and Brechling (1992)): instead it seems that taxes are shifted to the consumer by roughly one hundred percent, so tending to confirm what has long been a standard assumption in incidence analyses. The tentative finding of Keen (1998) that specific and ad valorem taxes on cigarettes have significantly different effects suggests that, in that market at least, quality and imperfect competition are phenomena of real importance.

Unemployment

Tax incidence analyses in the spirit of Harberger (1964) assume market clearing, and so cannot address tax effects operating through the level of unemployment. This precluded, until quite recently, consideration of an obviously important channel for public policy to have distributional consequences. Not before time, the area is now one of considerable activity, and one in which considerable progress has been made (by, amongst

[86]To see this, picture two curves showing utility as a function of income, one conditional on consuming the lower quality good the other conditional on consuming the higher. The point then follows from the observation that the latter cuts the former from below.

others, Atkinson (1994), Lockwood and Manning (1993) and Sørensen (1998)). Two results give a flavour of the consequences of explicitly modeling unemployment for assessing the distributional effects of public policy:

- Non-standard instruments—such as wage subsidies or a minimum wage— may become optimal (a point touched on in Section 4.4.3 above);
- Progressive taxation is good for employment. This result is robust against alternative specifications of the bargaining process (Koskela and Vilmunen (1996)) the underlying intuition being quite simple: greater progressivity makes increasing the post-tax wage more expensive in terms of the gross wage, tilting the balance in employee-firm negotiations towards higher employment rather than higher wages.

Incidence analysis in developing countries

The pervasiveness of distortions in some developing countries, although reduced by programs of structural adjustment, continue to raise particular issues for incidence analysis. For example, in the presence of quotas the incidence of a tariff may not be on consumers, even for a small country: instead the incidence may be on the holders of the quota rents. Thus a tariff even on a necessity could quite plausibly turn out to have a progressive effects, rather than a regressive one. Shah and Whalley (1991) provide many such examples arguing that presumptions on incidence conventional for developed countries may be quite inappropriate in the more distorted circumstances often found in developing ones.

5.1.4 The treatment of expenditure, and the contributory principle

The framework in Section 5.1.3 is deceptively simple. It is readily extended to non-linear taxes by working in terms of virtual prices and defining lump sum income to include virtual income. It deals neatly with public goods: by holding their provision constant, it avoids the need to allocate the benefits across households. In practice, however, much public expenditure is not of that form, and the question of how it should be treated in incidence studies is problematic.

Transfers are naturally thought of as simply negative taxes. But this may be inappropriate if they are perceived simply as a return to individuals from their past contributions. By the same token, payroll and other taxes that are perceived as buying entitlement to future benefits—pensions, health care, education—are not taxes in the normal sense of being unrequited by the government. Payments into actuarially fair social insurance schemes might thus be excluded from the analysis of redistribution. Implementation faces two difficulties. First, assessing the fairness of such

schemes is a difficult matter; this has been carefully addressed in the literature on intergenerational accounting addressed below. Second, it is unclear how to proceed if—as seems to be the case in the US, for example—contributors perceive the benefit entitlement purchased by their contributions to be bigger than it is. Practice differs in this area.

Tricky issues arise in connection with the provision by the government by goods that are largely private in nature, such as education, health care, some social services. These are generally either ignored, or their value imputed to recipients at cost: treated, that is, as a negative tax of that amount. In some cases, however, it may be hard to identify the true beneficiary of public provision. Take for example the case of education. To the extent that education increases future earnings ability, the beneficiary would appear to be the child, not the parent to whom, under standard procedures, the benefit would be allocated. In a Ricardian world, however, parents may be able to reap this gain for themselves by reducing their bequest to their children. This in turn may be easier for the rich, who plan to make bequests than for the poor who do not. Public provision may thus entitle the currently rich and the children of today's poor, a point that will not be picked up by the standard allocation procedure.

5.2 Fiscal Incidence Modelling Strategies

There are two broad approaches to measuring the incidence of fiscal systems—data-based and model-based approaches. In the data-based approach, one uses disaggregated data on the sources and uses of income by income class of household, and attributes the net burden of the various tax and transfer components to income class using assumptions about burden shifting. Neither the behaviour of households and firms nor relative price changes are taken explicitly into account, except through the shifting assumptions. The model-based approach constructs a general equilibrium model of the economy and calibrates it to actual data. It then lets the workings of the model determine the incidence pattern of various fiscal measures. This approach usually adopts a differential incidence approach to avoid having to deal with the general equilibrium effects of public expenditures, while the data-based approach can measure either absolute or differential incidence. The following summarizes the variants of these approaches.

The Tax-Transfer System and the Lorenz Curve

The statistical agencies of many countries routinely calculate tables of the size distribution of income. These show the proportion of the income earned by each of the five quintiles or ten deciles of the income distribution. For OECD countries, the pre-tax and transfer income distribution shows more than 40 percent of the income going to the top quintile, and less than 5 percent to the lowest quintile, a pattern which seems to be reasonably similar across countries and through time, although there is some

evidence of income becoming less equally distributed in recent years.

These tables, whose data refer to points on the Lorenz curve, are then adjusted in two ways to account for redistributive effects of government policies. First, transfer payments to households are added to income, and second, direct taxes are taken out. These changes typically show that direct personal taxes and transfers serve to even out income distributions to differing degrees in different countries.

These studies, although relatively straightforward to calculate and easy to interpret, are somewhat limited descriptions of fiscal incidence. Only direct taxes and transfer applying to households are included. And, the shifting assumptions are relatively crude: all taxes and transfers are assumed to be borne by those who are pay or receive them. It is not surprising that they typically show that the tax-transfer system is relatively progressive.

Detailed Data-Based Fiscal Incidence Studies

The shortcomings of the above approach are addressed in much more detailed fiscal incidence studies which include a much broader spectrum of taxes and transfers, and which use different incidence assumptions for the various tax sources. Standard studies using this approach include: Pechman and Okner (1974), Musgrave, Case and Leonard (1974) and Browning and Johnson (1979) for the United States; Nicholson and Britton (1976) and, annually, the Central Statistical Office for the United Kingdom; and Gillespie (1980) and Vermaeten *et al* (1995) for Canada. These studies tend to focus mainly on tax incidence and include the five main tax sources—the personal income tax, the corporate income tax, sales and excise taxes, payroll taxes, and property taxes. For each tax, the burden is allocated across income classes according to shifting assumptions which draw on standard theoretical and empirical analyses. It is commonly assumed that personal income and payroll taxes are borne by the income earners, sales taxes are borne by consumers and allocated according to consumption patterns across income groups, and corporate and property taxes tend to be allocated to capital earners, partly those in taxed sectors and partly capital income earners more generally. Results based on these types of assumptions typically show that the pattern of tax incidence is roughly proportional to income, a finding that belies the progressive tax as an important instrument in government's redistribution arsenal.

The proportionality result is a consequence of the mix of taxes and of the presumed incidence applying to each tax type. There has been some controversy in the literature about the assumptions adopted and their consequences for incidence. Some of the more important sources of controversy are as follows.

Perhaps the factor that most accounts for the absence of progressivity in the tax system is the assumption that sales and excise taxes are borne in proportion to consumption patterns out of income. Because low-income persons consume a much higher proportion of their incomes than do high-income persons, especially for highly taxed

goods, these taxes tend to be very regressive. Browning and Johnson (1979) have advocated that the incidence of sales and excise taxes ought to be based on factor incomes rather than consumption. They argue that because transfers to households are typically indexed for the price level, the effect of increases in the price level due to sales and excise taxes for low-income persons will be undone by compensating increases in transfers. When sales and excise taxes are allocated according to factor incomes, they turn out to be progressive rather than regressive.

Another reason why standard incidence assumptions might understate the true progressivity of the tax system is because of the way savings are treated. Sales and excise taxes are assumed to apply only to current consumption, and not to savings. Income taxes apply to savings to the extent that savings are done out of after-tax income. Once one views savings as future consumption, these incidence assumptions are more questionable. Sales and excise taxes do not apply to savings, but they do apply to the future consumption it yields. As well, savings generate capital income which itself will be subject to various taxes in the future. Taking all these into account will cause the tax system to appear more progressive than standard estimates would indicate.

The standard treatment of payroll taxes also contributes to the lack of measured progressivity. Not only do payroll tax rates apply just to earnings and not capital income, they also usually have an upper limit. But, it might be argued that payroll taxes should be omitted from incidence calculations because they are earmarked for particular social insurance programs that benefit contributors: any adverse effects on incidence should be offset by benefits received.

Whalley (1984), in a provocative piece, has shown that if incidence calculations are adjusted to take account of all the above arguments, the result will be a relatively progressive tax system rather than one which is merely proportional. By the same token, he has suggested a number of defensible adjustments that could be made in the opposite direction to make the tax system appear to be highly regressive. Most of these adjustments have to do with the interpretation of taxes on capital income. The standard analysis assumes that taxes on corporate income are borne by capital, either in the corporate sector or more generally. But, there are a number of reasons why this may not be appropriate. Corporations may be able to shift the burden of their taxes forward to consumers, or perhaps more likely backward to other factors, especially labour. This might be especially true in a small open economy. Taxes on personal capital income might also not be borne by asset owners, given the ease with which capital may be shifted abroad to avoid taxation, or shifted into untaxed assets (housing). It might even be argued that the part of labour income that is attributable to human capital accumulation will not be borne by the household, given that accumulating human capital is a substitute for accumulating assets with taxable capital income. Whalley shows that taking into account these considerations, the incidence of the tax system, in his case the Canadian one, can be made to appear very regressive. One is left with a very agnostic view of the true pattern of tax

incidence. Almost as agnostic, indeed, as Edwin Cannan was when expressing his doubts about the possibility of evaluating tax burdens to a Select Committee of the House of Commons over seventy years ago:

> I think that enquiry is a will-o'-the-wisp myself. I cannot help it if the House of Commons has asked you to do it.[87]

Whalley's results indicate just how sensitive fiscal incidence studies are to the shifting assumptions made. One could add other concerns to those expressed by Whalley. Incidence studies ignore the existence of unemployment and the policy responses to it, such as unemployment insurance. They also do not take account of the effects of non-tax instruments such as minimum wage laws, quantity controls in sectors like agriculture, and industrial subsidies, all of which have distributive effects. Perhaps most important, most studies concentrate largely on the tax side. As we have discussed above, it is becoming widely recognized that public expenditures are to a large extent devices for redistribution. Expenditure programs in areas of health, education, welfare and social insurance are ultimately motivated by redistributive goals. The extent to which these program succeed in redistributing real income remains elusive.

Some studies have included expenditures in their estimates of fiscal incidence, notably Gillespie (1980) for Canada, Le Grand (1982) for the United Kingdom and, more recently, van de Walle and Nead (1995) for developing countries (where public services form the bulk of redistribution policies, given the absence of comprehensive progressive tax systems). The methodology is analogous to that used to measure (absolute) tax incidence. Some assumption must be made about the incidence of the benefits of individual spending programs. For example, in studies in van de Walle and Nead, the actual costs of public services are allocated to income classes according to estimates of service usage. Obviously, this is a rather crude procedure, identifying as it does benefits with costs and ignoring any surplus that might be obtained by households. Nonetheless, the results are at least suggestive and in accordance with what one might expect. These studies find that public spending is at least mildly progressive in the sense that it is yields higher benefits for the poor when measured as a percentage of individuals' initial incomes, although it is regressive when it is measured in absolute value. But this broad overall incidence pattern masks important differences across public services. For example, public spending on primary and often secondary education is progressive, while it is regressive when spent on tertiary education. Likewise, primary health care centres are usually more pro-poor than hospital services, and in the former communist countries of Eastern Europe pension schemes are generally regressive while family allowances are progressive.

The expenditure side of the budget includes not only spending on goods and services but also transfers. Even though transfers are equivalent to negative taxes so might be

[87] Quoted by Prest (1955).

treated analogously to taxes, including them in incidence calculations is the exception rather than the rule. This may be partly because they are typically delivered as separate programs rather than through the tax system. In many cases, they are at least partly taken account of in defining the income base to include transfers received. As Whalley (1984) has shown, the tax system is much more progressive when measured as a proportion of income including transfers than when transfers are excluded. Clearly, the neglect of the expenditure side of the budget in fiscal incidence studies is a major shortcoming, despite the fact that public spending is an important part of redistributive policy. Indeed, its neglect is even more striking in model-based incidence studies discussed below.

A striking feature of most incidence studies is the apparently limited amount of redistribution they achieve. This might reflect both the theoretical limits to redistribution (Section 4) and the constraints imposed by the political system (Section 3). Goodin and Le Grand (1987) argue more generally that the non-poor inevitably benefit substantially from the welfare state. Using historical evidence from Australia, the United Kingdom and the United States, they argue that middle class support is needed to introduce major welfare state expenditures; that such programs will either be universal in focus, such as social insurance programs intended to alleviate insecurity, or if meant for the poor, will be infiltrated by the non-poor; and that, when under threat, welfare programs benefiting mainly the poor will be the most vulnerable. Indeed, for many social programs, the non-poor may benefit more than the poor. The non-poor are politically powerful so can manipulate the political agenda. They are better educated and able to exploit the administrative rules to their own ends. Their superior personal resources and flexible work schedules enable them to gain better access to public services. And, they are more able to make themselves appear to be poor so as to gain access to targeted programs. Goodin and Le Grand argue that inevitable middle class involvement in the welfare state is especially important in programs designed to achieve 'secondary' income redistribution, that is, to change the 'primary' income distribution obtained from participation in the market. They argue that a sensible strategy might instead be to focus redistribution policy on primary incomes by such instruments as employment policy, enhancing human capital investments, child care, the removal of barriers to employment by certain demographic groups, and even minimum wages. These forms of intervention might induce less involvement of the middle classes.

Microsimulations

The fabulously increased ease with which household-level micro-data can be manipulated has produced, from the 1980s, a different kind of micro-based study. This focusses not on the infrequent exercise of assessing the overall incidence of the tax system but on more routine evaluation of specific reforms: changes in the tax rate structure of income and payroll taxes for instance or in the tax treatment of the family. In this way it is possible to bring together detailed information on actual

households of the kind to be affected by the reform—moving beyond the hypothetical stylized case—with detailed information on often enormously complex tax-benefit structures.

Most applications of such methods continue to assume away behavioural effects. This is clearly problematic, not least because the very purpose of the reform may be to induce particular kinds of labour supply or other effects. Over this same period, equally dramatic developments have been made in the development of microeconometric methods designed precisely to recognize behavioural effects and incorporate them into the welfare analysis. Packages have been constructed that bring together the flexibility of micro-simulation analyses, able to handle a wide range of potential reforms, and the ability to incorporate behavioural responses. No doubt these capabilities will develop still further in the coming years.

These exercises continue to rest, however, on the same crude incidence assumptions as the incidence analyses described above; they largely neglect, that is, the general equilibrium consequences of the reforms they address.

Computable General Equilibrium Incidence Models

A major deficiency of the data-based fiscal incidence studies is their treatment of the interaction between the fiscal system and the market economy. By simply positing how various taxes are shifted, they do not explicitly take account of the complicated way in which relative prices and incomes might be influenced by distorting taxes. Computable general equilibrium models make the tax system a component of a fully-specified model of the market economy and allow incidence to be determined endogenously through the working of the economy. They thus avoid the need to assume rather arbitrary shifting assumptions.

Of course, this benefit is obtained at a cost. Computable, or applied, general equilibrium models may not be able to take advantage of the highly disaggregated data reflecting the complicated details of the fiscal system that are used in data-based incidence studies. Moreover, in place of the arbitrary shifting assumptions, one must select an almost equally arbitrary economic model along with parameters for technologies and tastes. One might argue that the choice of parameters can be made in a relatively informed way. Econometric estimates of elasticities of substitution in production and of systems of demand may be available. As well, the variables can be chosen so that the model can be calibrated to replicate the real world. And, sensitivity analysis can readily be carried out on the key variables in the model.[88]

Nonetheless, it is now well-known that the results of computable general equilibrium models can be very model-specific. This is illustrated by the work of Cox and Harris (1985) in their study of the effects of free trade on the Canadian economy. Contrary

[88]The use of computable general equilibrium models for policy analysis is outlined and defended in Shoven and Whalley (1984).

to the standard computable general equilibrium model with perfectly competitive industries operating with constant returns to scale, they treated the manufacturing industries as being imperfectly competitive, with economies of scale arising from fixed costs. They estimate the annual welfare gains from multilateral trade liberalization to be in the range of 8 to 10 percent of GNP, an order of magnitude greater than estimates obtained from models with perfect competition (which were generally less than 1 percent). Much of the gain is attributed to intra-industry rationalization to take advantage of economies of scale.

There is also an issue of interpreting the results of computable general equilibrium simulations. Incidence studies involve comparing the allocation achieved under the existing fiscal system with that of a benchmark case. The computation assumes that the existing capital stock and labour supply can be reallocated costlessly among alternative uses. Presumably the calculation cannot be thought to show what would, in fact, happen if the fiscal system were changed from the existing one to the benchmark case: the adjustment to a new equilibrium would take time, during which capital and labour supplies would have changed. The comparison must be viewed as hypothetical, comparing the economy as it is with what it might have looked like had a different tax system been in place some time ago. The fact that calculating the effects of moving from an existing general equilibrium allocation to an alternative one cannot be interpreted as describing what would actually happen if the tax system were to change makes it difficult to draw policy implications from the exercise.

The use of general equilibrium analysis to study the incidence of taxes had its genesis in the seminal work on the corporate tax by Harberger (1962), later synthesised and extended to other taxes by Mieszkowski (1969).[89] The original interest was in the effect of taxes on the functional distribution of income, essentially its effect on capital versus labour income. Harberger had argued that although the corporate tax was essentially a tax on the corporate sector of the economy, its incidence in general equilibrium fell capital in general, given the parameter values he used.

The use of computable general equilibrium models to study the incidence of taxes by income class has been quite limited. Most analyses have had more limited objectives, such as studying the effects of individual taxes or of particular tax reforms.[90] That reflects the difficulties in modelling economies to include the vast array of taxes in existence. One notable exception is Piggott and Whalley (1985), who compared the general equilibrium allocation of the UK tax system in 1973 with a benchmark system raising the same revenues with a proportional sales tax on all goods and services. In addition to finding relatively large excess burdens from the tax system (6–9 percent of

[89] This work studied the general equilibrium effects of differential changes in the tax system using comparative static analysis, and extrapolated it to discrete changes using linear approximations. Harberger's work was replicated by Shoven and Whalley (1972) using computable general equilibrium techniques and the results were found to be qualitatively virtually the same.

[90] A summary of several studies may be found in Shoven and Whalley (1984).

GDP, or almost one-quarter of government revenues), they find that the tax system is quite progressive, much more so data-based methodologies have generally suggested. In particular, the benchmark tax system leaves the top decile of the income distribution 25 percent worse of than the existing system, and the bottom decile 20 percent better off.

5.3 Intertemporal issues

5.3.1 Lifetime perspectives on tax incidence

The data-based fiscal incidence studies discussed above rely on annual income for classifying households. But, it is well-known that annual income distributions are heavily influenced by persons in extreme income positions for short periods of time. For example, lower income groups contain persons temporarily out of work as well as retired persons whose incomes might be much lower than their consumption. Higher income groups contain persons at the peaks of their earnings periods. It has been estimated that up to one half of inequalities in annual earnings can be attributed to variations of income over the life cycle (Lillard (1977); Blomqvist (1981)). That being the case, basing fiscal incidence on annual income statistics can be misleading. An alternative procedure is to calculate tax incidence on the basis of lifetime rather than annual incomes.

Different fiscal instruments will affect households differently in different periods of the life cycle. Sales and excise taxes are paid throughout the life cycle, while payroll taxes and property taxes are not. Income taxes also have an uneven lifetime profile, made more pronounced by their progressive nature. As well, transfers received tend to be concentrated in periods of low annual income. On a lifetime basis, they are likely to be much less progressive. The treatment of saving must also reassessed in a lifetime context. If households are life-cycle savers, sales and excise taxes are paid on both current consumption and savings and, as mentioned above, are likely to be less regressive on that account: Poterba (1989) finds, for example, that the distribution of excise payments in the US looks much less regressive when households are ordered by consumption than when they are ordered by income. But, if a high proportion of household savings is for bequests as some evidence suggests, sales and excise taxes might still be regressive on a lifetime basis. The overall impact of these life-cycle effects for fiscal incidence based on lifetime incomes are not obvious.

Measuring fiscal incidence on a lifetime basis is a very demanding task. Longitudinal lifetime income and tax profiles must be constructed on the basis of annual cross-sectional data, and that requires information on lifetime earnings, savings and inheritance and bequest patterns. An ambitious lifetime fiscal incidence study was undertaken by Davies, St-Hilaire and Whalley (1984) for the Canadian economy. Ap-

plying similar shifting assumptions to the five major tax types as above, they find that overall tax incidence is only mildly progressive, not unlike those found in annual studies. As well, the extent of progressivity varies relatively little as different shifting assumptions are adopted. These results reflect that fact that individual tax types which are progressive using annual data (e.g., income, corporate and property taxes) are less so using lifetime data, and the same for those which are regressive (sales and excise taxes and payroll taxes).

5.3.2 Intergenerational redistribution

Looking at fiscal incidence on a lifetime basis raises a further issue which has had considerable attention, and that is the effect that a fiscal program has on the lifetime real incomes of different age cohorts. As emphasised by Kotlikoff (1984), almost all fiscal measures have an intergenerational impact. The liabilities for income, consumption and payroll taxes differ systematically across the life cycle, so any change in the tax mix will induce intergenerational redistribution. For example, increasing sales taxes at the expense of payroll taxes will redistribute from older to younger cohorts. Transfers tend to occur especially in old age, implying that public pension schemes that are financed out of current taxes will tend to redistribute from younger to older cohorts. Expenditures on public goods and services also have obvious age-specific benefits. And, deficit financing by the government is equivalent to a transfer from the future to current taxpayers.

Merely identifying the generational impact of these various policy measures is not sufficient to indicate the extent of intergenerational redistribution that ultimately occurs. Two mitigating effects must be considered. First, intergenerational redistribution imposed by the government may be partially offset by changes in private intergenerational transfers (bequests). Though the theoretical basis for full Ricardian neutrality has been considerably undermined by the critique of Bernheim and Bagwell (1988) as discussed earlier, nonetheless saving for future generations may have some characteristics of the voluntary provision of a public good for which the Warr (1983) neutrality theorem applies. Moreover, to the extent that bequests are involuntary and related to the accumulation of wealth for precautionary purposes, some offsetting of public intergenerational transfers might occur. Second, net intergenerational redistribution is considerably dampened in the long run when each cohort is a recipient from previous generations as well as a transferer to following ones. In the steady state, the only net redistribution that occurs is that which is made possible by growth in per capita incomes. In other words, most of the intergenerational redistribution that occurs happens during the transition. The implications is that estimates of intergenerational incidence may need to be done over relatively long periods of time.

Two lines of research into the intergenerational impact of budgetary policies have been pursued, corresponding to the model-based and the data-based approaches to fiscal

incidence described above. In each case, the focus has been mainly on identifying intergenerational redistributive effects rather than intragenerational ones.

In the first, dynamic overlapping-generations models are used to simulate the effect of fiscal policy changes that reallocate tax burdens across generations. In the seminal work, Summers (1981) constructed a simple single-sector neo-classical growth model with continuously overlapping generations of representative households. Households of different cohorts supplied identical quantities of labour over their fixed working lives and decided only how to reallocate it to a lifetime consumption profile using capital markets. Wage rates rose proportionately over time due to technical progress. In this setting, Summers simulated the steady state effects of replacing the existing system of taxes on wages and capital income in the US with a tax on wages alone, and with a tax on consumption alone, in both cases to finance a given stream of government expenditures. Removing the tax on capital income would eliminate the only distortion in the tax system, and presumably stimulate savings. The two tax substitutions would also affect savings by affecting a pure intergenerational transfer. Thus, since a wage tax imposes tax liabilities earlier in the life-cycle than a combined wage and capital income tax, substituting the former for the latter amounts to an ongoing intergenerational transfer from the young to the old, analogous to unfunded public pensions. This tends to reduce savings. On the other hand, substituting a consumption tax for an income tax will have the opposite effect of transferring from the old to the young, increasing savings. Summers obtained rather large effects from his simulations. In his base-case scenario, the wage tax substitution increased the capital-output ratio by 42 percent, per capita lifetime consumption by 13 percent and lifetime utility by 5 percent in the steady state. The consumption tax effects were even larger: the capital-output ratio rose by 54 percent, per capita lifetime consumption by 16 percent, and lifetime utility by 12 percent.

In retrospect, the source of these dramatic increases was apparent. In Summers' economy, the capital stock was well below its 'Golden Rule' optimum. That is, the steady-state rate of interest was 10.5 percent compared with a rate of growth in income of 3.5 percent: the implicit return on an additional pound of investment was well in excess of the implicit rate of return on an additional pound of intergenerational transfer from the young to the old, so households would be made better off by reducing the latter and stimulating savings instead. And, since the amount of stimulation of savings was that much greater for the consumption tax than for the wage tax, the welfare effects were correspondingly larger.

Summers' work was influential both in academic work and policy discussion, for it suggested that a move away from capital income taxation would reap immeasurable benefits. But, the approach was not without problems. For one things, some of the assumptions built into the model tended to bias the simulated effects on savings. These included the absence of a labour-leisure choice, which would make both a consumption and a wage tax distortionary; the assumption of no bequests, either intentional

or unintentional; the assumption of perfect capital markets, especially the absence of liquidity constraints which would imply that increased wage taxes might reduce consumption rather than savings; the assumption of an exponentially increasing wage profile over the life cycle; the assumption that taxes were proportional rather than progressive; the assumption of an exogenous retirement date and length of life; and the assumption of perfect foresight among households.

Much of the subsequent literature addressed these issues,[91] but one further issue was of overriding importance—the restriction of simulated intergenerational redistributive effects to steady-state ones. Much of the welfare gain in the steady state can be attributed to an increase in the capital stock. Since the adjustment to a higher capital stock entails forgoing current consumption, those generations alive while the bulk of the adjustment takes are liable to suffer losses. To the extent that is true, some of the long-run welfare gains will be offset by short-term welfare losses. Indeed, if the increased savings is due to an intergenerational transfer rather than an elimination of the distortion on capital markets, the gain to generations in the long run is simply a redistribution from current generations and cannot be viewed as an efficiency gain. Thus, simulating the temporal time pattern of intergenerational redistributive effects would seem to be mandatory. This is precisely what has been reported in Auerbach and Kotlikoff (1987).

Auerbach and Kotlikoff construct an dynamic overlapping-generations model in which all agents have perfect foresight, and use it to simulate the effects of tax substitutions on the time path of resource allocation beginning from an initial steady state, and on the welfare of all cohorts along the path. Though similar in structure to Summers' model, their model also addresses some of the shortcomings mentioned above: labour supply and retirement ages are endogenous, the age-earnings profile is strictly concave, and taxes are progressive.[92] As well, the policy simulations include not only tax substitutions analogous to those of Summers, but other fiscal policies with intergenerational impacts, such as public pensions and deficit finance. Not surprisingly, in the case of the tax reform simulations, tracing out both the short-run and long-run consequences of the tax substitutions considerably tempers the impact of the Summers results. Switching from an income tax to a wage tax still produces some increase in the capital stock in the long run, and to a consumption tax a larger increase. But, the transitional analysis reveals that not all age cohorts share in the benefits of this. In the case of the consumption tax, cohorts who are young during the transition as well as all future cohorts obtain gains in lifetime utility, but older cohorts are made worse off. When a wage tax is substituted for an income tax, the effects are the opposite. Younger and future generations are made worse off by the change, despite the fact that the capital stock has risen. (With variable labour supply there is now a

[91] See the summary in Boadway and Wildasin (1994).

[92] One shortcoming that Auerbach and Kotlikoff share with Summers is the neglect of intragenerational effects. In both studies, all households of a given cohort are identical.

distortion introduced by wage taxation.) Older generations are made better off.

The companion simulations that Auerbach and Kotlikoff conduct on changes in public pensions and deficits show comparable effects. An increase in unfunded public pensions entails an intergenerational transfer from the young to the old which is analogous to, say, a substitution of a wage tax for a consumption tax. And, both are analogous to the substitution of deficit for tax financing of government expenditures. This leads them to the view that the budget deficit as conventionally reported by governments is a misleading indicator of the impact of the budget structure on the economy and on the well-being of various generations. The implicit liability created by unfunded social security or by tax reforms that change the timing of tax liabilities over the life cycle are indistinguishable in their economic effects from the budget deficit. This has led the authors to propose an alternative form of reporting, on which stresses the full impact of budgetary policies on different generations—*generational accounting*.[93]

Like the above simulation models, generational accounting focuses exclusively on the effect of fiscal policies on representative members of different generations. But, as with data-based incidence studies, it does so by assigning net benefits to various cohorts using some presumed notion of burden-bearing without taking account of behavioural or relative price changes. The methodology is straightforward, and is based upon the intertemporal budget constraint unavoidably faced by government:

$$\sum_{s=0}^{T} A_{t,t-s} + \sum_{s=1}^{\infty} A_{t,t+s} = \sum_{s=t}^{\infty} \frac{G_s}{(1+r)^{t-s}} - W_t$$

where $A_{t,k}$ is the present value at time t of the remaining net tax payments of persons in cohort k (its 'generational account'), G_s is government expenditures on goods and services in year s, W_t is the stock of government wealth in year t, and r is the discount rate, assumed to be constant. The first term on the lefthand side reflects the generational accounts of all cohorts currently living, while the second is for future generations. Implementing the notion of generational accounting involves calculating for some presumed future path of budgetary policies the values of $A_{t,k}$ and G_s that satisfy this budget constraint. Auerbach et al (1991) propose calculating generational accounts for the case where current policies, both tax-transfer policies and expenditure policies, are kept intact into the future. Based on these policies, and using forecasts of population and average taxes and transfers for each cohort, generational accounts are calculate for all cohorts currently live. Next, values for the righthand side terms are calculated, again based on projections of current policies into the indefinite future. Then, the sum of generational accounts for future generations is obtained simply from the budget deficit. Finally, the average lifetime net tax burden to be faced by future generations is calculated, assuming that the net tax payment rises at the rate of productivity growth for the economy (which itself must be assumed). In effect, the

[93]See Auerbach et al (1991) for the initial contribution. Subsequent summaries may be found in Kotlikoff (1992) and Auerbach et al (1994). See also the critical view of Haveman (1994).

net liabilities of the government as of today are amortized over all future generations: they all share in the paying debt.

Auerbach et al (1994) report generational accounts as of 1991 for males and females separately. They find that, among the currently alive generations, older generations (those 60 and over) will have negative generational accounts: they will be net recipients of transfers. Younger generation will have positive accounts, with the peak being for those aged 30 for males and 25 for females. Males have systematically larger generational accounts than females of similar age. Significantly, representative members of future generations have much higher generational accounts than the youngest member of the current generation. In the case of both males and females, the generational accounts of future cohorts are over twice as high as those of the current newborns, reflecting what they refer to as a current bias in government budgets.

Generational accounts are clearly in their infancy, and the magnitudes reported by Auerbach et al are likely to be controversial. They are based on projections that assume that current policies are likely to remain in effect forever, but that the debt will be spread equally over all generations yet to be borne. Nonetheless, they do serve to complement existing budgets and incidence estimates by highlighting the intergenerational biases built into existing policies.

5.4 Explaining redistribution

There has been strikingly little empirical work seeking to explain observed patterns of redistribution. But the list is growing, and some instructive stylised facts beginning to emerge, albeit in some cases still tentatively. Here we simply indicate some of the issues and broad findings.

- Redistribution tends to be greater the wider the franchise: see, for instance, Metzler and Richard (1981) and Lindert (1994, 1996). This is doubtless as one would expect, though quite why the initial elite should decide to widen the electorate may not always be as obvious as it seems.[94]

- Demographics play a powerful role. Lindert (1994, 1996) shows how the development of pension schemes, for example, was related to greying of the population.

- Attention has recently been drawn to a strong positive correlation between openness and government size. Rodrik (1998) attributes this to the greater vulnerability of open economies to uninsured shocks, creating a greater social insurance

[94] There has been surprisingly little theoretical work on the role of the franchise and redistribution, and this has been reflected in omission of the topic from this survey. Such work is beginning to emerge however: see Acemoglu and Robinson (1997).

role for government. Alesina and Wacziarg (1998) argue instead that the correlation reflects rather a positive correlation between openness and size, with scale economies causing the share of government to be lower in large economies. Or the correlation might reflect the administrative advantages that openness offers in levying taxes.[95]

- Attitudinal data suggests a considerable dose of self-interest in the support for redistribution. There is evidence, for example, that public health care is supported most strongly by those likely to become recipients, as for unemployment benefit too.

- Perhaps most intriguingly, there are recurring signs that redistribution tends to be greater the *less* underlying pre-tax inequality there is. First noted by Peltzman (1980), the effect is noted too by Bénabou (1997), Lecallion et al (1984), Persson (1995) and emphasized in this volume by Lindert (1998), who dubs this the 'Robin Hood paradox.' At a more introspective level, the combination in both the US and UK of increased inequality and reduced political support for redistribution suggests the same pattern.

  Causation is evidently a concern here. Greater redistribution leading to paradoxically more inequality could be explained in terms of risk-taking induced by social insurance, along the lines discussed in Section 2.2.2. As an empirical regularity, however, redistribution does appear to be negatively related to prior inequality. Such a link is not easily explained. It runs exactly counter to the implication of the standard models of linear taxation described above. An increase in inequality would be expected to reduce the median wage relative to the mean, so leading to more progressivity in the voting model of Section 3.1.2. In the optimal linear income tax model of Section 4.3, one would again expect an increase in pre-tax inequality to be associated with more redistribution, not less.

  It is not entirely easy to see how the link might run the other way. Persson (1995) attempts an explanation in terms of interdependent preferences. Peltzman's explanation runs in terms of diminished commonality between middle and lower income groups as inequality increases. Perhaps the explanation—building on the importance of self-interest in motivating redistribution—is that it is not increased inequality at any moment which matters but rather the extent to which low spells of low income are concentrated amongst particular groups. The more concentrated such spells are, the less inclined those least subject to such spells will be to support schemes that would protect them if they did fall on hard times.

[95] While Rodrik addresses and rejects this, the administrative advantages will extend beyond the trade taxes that he focusses on (extending, not least, to VAT).

6 Conclusions

Such a long paper should not be burdened with a long conclusion. There are, however, a few themes that merit some emphasis and comment.

- The recent literature has shown that policy instruments widely criticized as being inefficient—such as in-kind welfare support, production subsidies, minimum wages—may have a proper place in the government's armory of redistributive tools. They may enable a weakening of the self-selection constraints that limit the amount of redistribution which can be achieved, and they may restrict the damage that ill-intentioned policymakers can do.

  These arguments, however, must be applied with care. Just as the results of the new trade literature, establishing potential benefits from trade restrictions, are recognized to hold the potential for abuse, so too may these justifications for non-standard policy instruments. Some private interests are evidently served, for example, by arguments for public provision of private goods, or even for a minimum wage.

- The politics of redistribution are complex. Empirical work has given rise to Director's' Law, that much redistribution is to the centre; and Disraelian conservatism in the UK has long built on the supposed natural conservatism of the working poor. Recent theoretical analyses have begun to explain why it might indeed be that high and low income roups find themself allied against middle groups in the public provision of private good: the poor would rather do without, the rich would rather fend for themselves.

  Understanding of these and related positive issues—such as the widespread perception that it is easier to sustain political support for universal rather than means-tested programs—is hampered by the lack of progress in understanding the technically challenging issue of voting over non-linear tax schedules.

- It is still the case that very little is known about the effective incidence of some—perhaps most—key taxes. Understanding of the distributional impact of, for example, the corporate tax, dividend tax and capital gains tax has not progressed greatly (from a low base) over the last twenty years or so. Much policy advice continues to be based on naive incidence assumptions: that major welfare reforms will leave the level and distribution of pre-tax wages unchanged, for instance. Moreover, theoretical developments over this period have expanded the range of our ignorance by showing that the competitive analyses previously developed do not even bound the possible outcomes under when markets are imperfect. More positively, advances in the understanding of imperfect labour markets provide a real gain in the applicability of incidence analysis, and it may be that the empirical work that these theoretical advances have started to

foster will generate real advances on the perennially troublesome issue of effective incidence.

- Much empirical work on redistribution over the last two decades or so has focused on micro data, exploiting the enormous technological advances over the period. Recent work from a more macro perspective, however, has raised a range of intriguing issues.

  One such is the question of how redistribution affects growth. Another is the apparent regularity that more equal societies apparently redistribute more (an observation that also seems consistent with the casual observation that increased inequality in the UK and US during the 1980s has been associated with a turn of political tide against redistributive programs). This runs exactly counter to the predictions of our standard models. Finding a coherent explanation—which may require exploiting micro-level data—may bring a deeper understanding of the underlying rationale for redistributive programs.

- The relationship between equity and efficiency considerations is much more complex than the traditional view of an unavoidable trade-off. Capital market and other imperfections provide quite coherent reasons to suppose that redistribution may in some circumstances increase growth: equity and efficiency may go together. Happy though such an outcome is, eyes should not be closed to the likelihood that, beyond some point, the two will indeed conflict. The range of conflict between equity an efficiency may be narrower than previously thought; but it has not gone away. And it is ironic that economists have produced reasons to suppose that redistribution may be good for growth at precisely the time that voters in so many countries have been persuaded of exactly the opposite.

References

Aaron, A.A. (1966), The social insurance paradox, Canadian Journal of Economics and Political Science 32: 371-376.

Abel, A., G. Mankiw, L. Summers and R. Zeckhauser, Assessing dynamic efficiency: Theory and evidence, Review of Economic Studies, 56: 1-20.

Acemoglu, D. and J. Robinson (1997), Why did the West extend the franchise?: Democracy, inequalty and growth in historical perspective, MIT Working Paper No. 97-23.

Aghion, P. and P. Bolton (1997), A theory of trickle-down growth and development, Review of Economic Studies, 64: 151-172.

Ahmad, E. and N. Stern (1984), The theory of tax reform and Indian indirect taxes, Journal of Public Economics, 25: 259-298.

Ahmad, E. and N. Stern (1991), The Theory and Practice of Tax Reform in Developing Countries (Cambridge University Press, Cambridge, UK).

Akerlof, G.A. (1978), The economics of 'tagging' as applied to the optimal income tax, welfare programs, and manpower training, American Economic Review, 68: 8-19.

Alesina, A. (1988), Credibility and policy convergence in a two-party system with rational voters, American Economic Review, 78: 796-805.

Alesina, A. and R. Wacziarg (1998), Openness, country size and government, Journal of Public Economics, 69: 305-321.

Alesina, A. and D. Rodrik (1994), Distributive politics and economic growth, Quarterly Journal of Economics, 109: 465-490.

Alesina, A. and H. Rosenthal (1995), Partisan Politics, Divided Government and the Economy (Cambridge University Press; Cambridge, UK).

Allen, F. (1982), Optimal linear income taxation with general equilibrium effects on wages, Journal of Public Economics, 17: 135-143.

Allingham, M. (1979), Inequality and progressive taxation: An example, Journal of Public Economics, 11: 273-274.

Andreoni, J. (1990), Impure altruism and donations to private goods; A theory of warm-glow giving, Economic Journal, 100: 464-477.

Arrow, K.J, (1951), Social Choice and Individual Values (Yale University Press; New Haven, Conn.).

Arrow, K.J. (1951), An extension of the basic theorems of classical welfare economics, in: Proceedings of the Second Berkeley Symposium (Berkeley: University of California Press) 507-532.

Arrow, K.J. (1963), Uncertainty and the welfare economics of medical care, American Economic review, 53: 942-973.

Atkinson, A.B. (1973), How progressive should income tax be, in: M. Parkin and A.R. Nobay, eds., Essays in Modern Economics (Longman, London).

Atkinson, A.B. (1991), Social insurance, Geneva Papers on Risk and Insurance, 16: 113-131.

Atkinson, A.B. (1994), The distribution of the tax burden, in: J.M. Quigley and E. Smolensky, eds., Modern Public Finance (Harvard University Press: Cambridge, Mass.) 13-49.

Atkinson, A.B. and N. Stern (1974), Pigou, taxation and public goods, Review of Economic Studies, 41: 119-128.

Atkinson, A.B. and J.E. Stiglitz (1972), The structure of indirect taxes and economic efficiency, Journal of Public Economics, 1: 97-119.

Atkinson, A.B. and J.E. Stiglitz (1976), The design of tax structure: direct versus indirect taxation, Journal of Public Economics, 6: 55-75.

Atkinson, A.B. and J.E. Stiglitz (1980), Lectures in Public Economics (McGraw-Hill, Maidenhead).

Auerbach, A.J. and L.J. Kotlikoff (1987), Dynamic Fiscal Policy (Cambridge University Press, Cambridge, U.K.).

Auerbach, A.J., J. Gokhale and L.J. Kotlikoff (1991), Generational accounting: a meaningful alternative to deficit accounting, in: D. Bradford ed., Tax Policy and the Economy, vol. 5 (MIT Press, Cambridge, Mass.) 55-110.

Auerbach, A.J., J. Gokhale and L.J. Kotlikoff (1994), Generational accounting: a meaningful way to evaluate fiscal policy, Journal of Economic Perspectives, 8: 73-94.

Aumann, R.J. and M. Kurz (1977), Power and taxes, Econometrica, 45: 1137-1161.

Baker, P. and V. Brechling (1992), The impact of excise duty changes on retail prices in the UK, Fiscal Studies, 13: 48-65.

Balasko, Y. and K. Shell (1980), The overlapping generations model, I: the case of pure exchange without money, Journal of Economic Theory, 23:281-306.

Banerjee, A.V. and A.F. Newman (1993), Occupational choice and the process of development, Journal of Political Economy, 101: 274-298.

Barr, N. (1993), The Economics of the Welfare State (Oxford University Press, Oxford).

Barro, R.J. (1974), Are government bonds net wealth?, Journal of Political Economy, 82: 1095-1117.

Beaudry, P. and C. Blackorby (1997), Taxes and employment subsidies in optimal redistribution programs, Discussion Paper No. 97-21, Department of Economics, University of British Columbia, Vancouver, Canada.

Becker, G.S. (1968), Crime and punishment: An economic approach, Journal of Political Economy, 76: 169-217.

Becker, G.S (1981), A Treatise on the Family (Harvard University Press, Cambridge, U.S.).

Becker, G.S. (1983), A theory of competition among pressure groups for political influence, Journal of Political Economy, XCVIII: 371-400.

Bénabou, R. (1996), Inequality and growth, NBER Macroeconomics Annual (MIT Press, Cambridge, Mass.) 11-74.

Bergstrom, T.C., L. Blume and H. Varian (1986), On the private provision of public goods, Journal of Public Economics, 29: 25-49.

Bernheim, B.D. and K. Bagwell (1988), Is everything neutral?, Journal of Political Economy, 96: 308-338.

Bertola, G. (1993), Factor shares and savings in endogenous growth, American Economic Review, 83: 1184-1198.

Bertola, G. (1998), Macroeconomics of distribution and growth, this volume, chapter 9.

Besley, T. and A. Case (1995), Does electoral accountability affect economic policy choices? Evidence from gubernatorial term limits, Quarterly Journal of Economics, 110: 769-796.

Besley, T. and S. Coate (1992), Workfare versus welfare: incentive arguments for work requirements in poverty-alleviation programs, American Economic Review, 82: 249-261.

Besley, T. and S. Coate (1997), An economic model of representative democracy, Quarterly Journal of Economics, 112: 85-114.

Besley, T. and J. McLaren (1993), Taxes and bribery: The role of wage incentives, Economic Journal, 103: 119-141.

Besley, T. and H. Rosen (1994), Sales taxes and prices: An empirical analysis, NBER Working Paper No. 6667.

Bhagwati, J.N., R. Brecher and T. Hatta (1983), The generalized theory of transfers and welfare: Bilateral transfers in a multilateral world, American Economic Review, 73: 606-618.

Black, D. (1948), On the rationale of group decision-making, Journal of Political Economy, 56: 23-34.

Blake, R. (1966), Disraeli (Methuen, London).

Blackorby, C. and D. Donaldson (1988), Cash versus kind, self selection and efficient transfers, American Economic Review, 78: 691-700.

Blomquist, S. and V. Christiansen (1995), Public provision of private goods as a redistributive device in an optimum income tax model, Scandinavian Journal of Economics, 97: 547-567.

Blomqvist, N.S. (1981), A comparison of distributions of annual and lifetime income: Sweden around 1970, Review of Income and Wealth 27: 243-264.

Boadway, R. and N. Bruce (1984), Welfare Economics (Basil Blackwell, Oxford).

Boadway, R.W. and F.R. Flatters (1982), Efficiency and equalization payments in a federal system of government: a synthesis and extension of recent results, Canadian Journal of Economics, 15: 613-633.

Boadway, R. and M. Keen (1993), Public goods, self-selection and optimal income taxation, International Economic Review, 34: 463-478.

Boadway, R. and M. Keen (1998), Evasion and time consistency in the taxation of capital income, International Economic Review, 39: 461-476.

Boadway, R. and N. Marceau (1994), Time consistency as a rationale for public unemployment insurance, International Tax and Public Finance, 1: 107-126.

Boadway, R. and M. Marchand (1995), The use of public expenditures for redistributive purposes, Oxford Economic Papers, 47: 45-59.

Boadway, R., N. Marceau, and M. Marchand (1996), Investment in education and the time inconsistency of redistributive tax policy, Economica, 63: 171-189.

Boadway, R., N. Marceau and M. Sato (1999), Agency and the design of welfare systems, Journal of Public Economics, forthcoming.

Boadway, R., M. Marchand and M. Sato (1998), Subsidies versus public provision of private goods as instruments for redistribution, Scandinavian Journal of Economics, 100: 545-564.

Boadway, R., M. Marchand and M. Vigneault (1998), The consequences of overlapping tax bases for redistribution and public spending in a federation, Journal of Public Economics, 68: 453-478.

Boadway, R.W. and D.E. Wildasin (1989), A median voter model of social security, International Economic Review, 30: 307-328.

Boadway, R. and D. Wildasin (1994), Taxation and savings: A survey, Fiscal Studies, 15: 19-63.

Bordignon, M. (1993), A fairness approach to income tax evasion, Journal of Public Economics, 52: 345-362.

Boskin, M.J. (1973), Local tax and product competition an the optimal provision of public goods, Journal of Political Economy, 81: 203-210.

Bourguignon, F. and P.-A. Chiappori (1992). Collective models of household behaviour: An introduction, European Economic Review, 36: 355-364.

Brennan, G. and J. Buchanan (1980), The Power to Tax: Analytical Foundations of a Fiscal Constitution (Cambridge University Press, Cambridge, UK).

Brennan, G. and J. Buchanan (1977), Towards a constitution for Leviathan, Journal of Public Economics, 8: 255-273.

Breton, A. (1993), Toward a presumption of efficiency in politics, Public Choice, 77: 53-65.

Brett, C. (1998), Who should be on workfare? the use of work requirements as part of an optimal tax mix, Oxford Economic Papers, 50: 607-622.

Browning, E.K. (1975), Why the social insurance budget is too large in a democracy, Economic Inquiry, 13: 373-388.

Browning, E. (1976), The marginal cost of public funds, Journal of Political Economy, 84: 283-298.

Browning, E.K. and W.R. Johnson (1979), The Distribution of the Tax Burden (The American Enterprise Institute, Washington).

Bruce, N. and M. Waldman (1991), Transfers in kind: why they can be efficient and nonpaternalistic, American Economic Review, 81: 1345-1351.

Bruno, M., M. Ravallion and L. Squire (1998), Equity and growth in developing countries: Old and new perspectives on the policy issues, in: V. Tanzi and K-Y. Chu,

eds., Income Distribution and High Quality Growth (MIT Press, Cambridge, Mass.).

Buchanan, J.M. (1980), Rent seeking and profit seeking, in: J.M. Buchanan, R.D. Tollison and G. Tullock, eds., Toward a Theory of the Rent-Seeking Society (Texas A&M Press, College Station, Texas), 3-15.

Buchanan, J.M. and G. Tullock (1962), The Calculus of Consent (The University of Michigan Press, Ann Arbor, Michigan).

Bucovetsky, S. (1991), Choosing tax rates and public expenditure levels using majority rule, Journal of Public Economics, 46: 113-131.

Calvo, G.A. and M. Obstfeld (1988), Optimal time-consistent fiscal policy with finite lifetimes, Econometrica, 56: 411-432.

Chandar, P. and L.L. Wilde (1998), A general characterization of optimal income tax enforcement, Review of Economic Studies, 65: 165-183.

Clark, D. and J. Thomas (1995), Probabilistic voting, campaign contributions, and efficiency, American Economic Review, 85: 254-259.

Coate, S. (1995), Altruism, the samaritan's dilemma, and government transfer policy, American Economic Review, 85: 46-57.

Coate, S. and S. Morris (1995), On the form of transfers to special interests, Journal of Political Economy, 103: 1210-1235.

Corlett, W.J. and D.C. Hague (1953), Complementarity and the excess burden of taxation, Review of Economic Studies, 21: 21-30.

Coughlin, P. (1984), Expectations about voter choices, Public Choice, 44: 49-59.

Coughlin , P. (1986), Elections and income distribution, Public Choice, 50: 27-91.

Cournot, A. (1960, originally published 1838), Researches into the Mathematical Principles of the Theory of Wealth (Frank Cass & Co., London).

Cowell, F.A. (1990), Cheating the Government: The Economics of Evasion (MIT Press, Cambridge, Mass.).

Cox, D. and R. Harris (1985), Trade liberalization and industrial organization: some estimates for Canada, Journal of Political Economy, 93: 115-145.

Cremer, H., V. Fourgeaud, M. Leite-Monteiro, M. Marchand and P. Pestieau (1996), Mobility and redistribution: A survey, Public Finance 51: 325-352.

Cremer, H. and F. Gahvari (1996), Tax evasion and optimal general income tax, Journal of Public Economics, 60: 235-249.

Cremer, H., M. Marchand and P. Pestieau (1990), Evasing, auditing and taxing: The equity-equity-compliance tradeoff, Journal of Public Economics 43: 67-92

Dahlby, B.G. (1981), Adverse selection and Pareto improvements through compulsory insurance, Public Choice, 37: 547-558.

Dasgupta, P. and P. Hammond (1980), Fully progressive taxation, Journal of Public Economics, 13: 141-154.

Dasgupta, P. and D. Ray (1986), Inequality as a determinant of malnutrition and unemployment: Theory, Economic Journal, 96: 1011-1034.

Dasgupta, P. and D. Ray (1987), Inequality as a determinant of malnutrition and unemployment: Policy, Economic Journal, 97: 177-188.

Dasgupta, P. A.K. Sen and D. Starrett (1973), Notes on the measurement of inequality, Journal of Economic Theory, 6: 180-187.

Davies, J., F. St-Hilaire and J. Whalley (1984), Some calculations of lifetime incidence, American Economic Review, 74: 633-649.

de Meza, D. and D.C. Webb (1987), Too much investment: A problem of asymmetric information, Quarterly Journal of Economics, 102: 281-292.

Deaton, A. and N. Stern (1986), Optimally uniform commodity taxes, taste differences and lump-sum grants, Economics Letters, 20: 263-266.

Dhami, S. (1997), Political economy of tax policy under asymmetric information, Mimeo, University of Toronto.

Diamond, P.A. (1975), A many-person Ramsey tax rule, Journal of Public Economics, 4: 335-342.

Diamond, P. (1998), Optimal income taxation: An example with a U-shaped pattern of optimal marginal tax rates, American Economic Review, 88: 83-95.

Diamond, P.A. and J.A. Mirrlees (1971), Optimal taxation and public production II: Tax rules, American Economic Review, 61: 261-278.

Diamond, P. and E. Sheshinski (1995), Economic aspects of optimal disability benefits, Journal of Public Economics, 57: 1-23.

Dilnot, A., J.A. Kay and M.J. Keen (1990), Allocating taxes to households: A methodology, Oxford Economic Papers, 42: 210-230.

Dixit, A.K. (1996), The Making of Economic Policy (MIT Press, Cambridge, Mass.).

Dixit, A.K., G.M. Grossman and E. Helpman (1996), Common agency and coordination: General theory and application to government policy-making, Journal of Political Economy, 105: 752-769.

Dixit, A.K. and J. Londregan (1995), Redistributive politics and economic efficiency, American Political Science Review, 89: 856-866.

Dixit, A.K. and J. Londregan (1996), The determinants of success of special interests in redistributive politics, Journal of Politics, 58: 1132-1155.

Dixit, A.K. and J. Londregan (1998), Ideology, tactics, and efficiency in redistributive politics, Quarterly Journal of Economics, 124: 497-529.

Dixit, A.K. and V. Norman (1980), Theory of International trade (Cambridge University Press; Cambridge, Mass.).

Downs, A. (1957), An Economic Theory of Democracy (Harper and Row, New York).

Drèze, J. and N. Stern (1987), The Theory of Cost-Benefit Analysis, in: A.J. Auerbach and M. Feldstein, eds., Handbook of Public Economics, Volume II, (North Holland, Amsterdam), 909-989.

Eaton, J. and H.S. Rosen (1980), Optimal redistributive taxation and uncertainty, Quarterly Journal of Economics, 95: 357-364.

Ebert, U. (1992), A reexamination of the optimal nonlinear income tax, Journal of Public Economics, 49: 47-73.

Edgeworth, F.Y. (1881), Mathematical Psychics (Routledge and Kegan Paul, London).

Edwards, J., M. Keen and M. Tuomala (1994), Income tax, commodity taxes and public good provision: a brief guide, Finanzarchiv, 51: 472-497.

Epple, D. and R.E. Romano (1996a), Ends against the middle: Determining public service provision when there are private alternatives, Journal of Public Economics, 62: 297-325.

Epple, D. and R.E. Romano (1996b), Public provision of private goods, Journal of Political Economy, 104: 57-84.

Fair, R.C. (1971), The optimal distribution of income, Quarterly Journal of Economics, 85: 551-579.

Feldstein, M.S. (1972), Distributional equity and the optimal structure of public prices, American Economic Review, 62: 32-36.

Fellman, J. (1976), The effects of transformations on Lorenz curves, Econometrica, 44: 869-881.

Fischer, S. (1980), Dynamic inconsistency, cooperation and the benevolent dissembling government, Journal of Economic Dynamics and Control, 2: 93-107.

Flatters, F. and W.B MacLeod, (1995), Administrative corruption and taxation, International Tax and Public Finance, 2: 397-417.

Fleurbaey, M. and Maniquet, F. (1998), Optimal income taxation: An ordinal approach, CORE, Université Catholique de Louvain, mimeo.

Fogel, R.W. and S.L. Engerman (1971), Time on the Cross: The Economics of American Negro Slavery (Little, Brown and Co., Boston).

Frank, R.H. (1987), If *homo economicus* could choose his own utility function, would he want one with a conscience?, American Economic Review, 77: 593-604.

Galor, O. and J. Zeira (1993), Income distribution and macroeconomics, Review of Economic Studies, 60: 35-52.

Gans, J.S. and M. Smart (1996), Majority voting with single-crossing preferences, Journal of Public Economics, 59: 219-237.

Gillespie, W.I. (1980), The Redistribution of Income in Canada (Gage Publishing Ltd., Ottawa).

Goodin, R.E. and J. Le Grand (1987), Not Only the Poor: The Middle Classes and the Welfare State (Allen and Unwin, London).

Gordon, J.P.F. (1989), Individual morality and reputation costs as deterrents to tax evasion, European Economic Review, 33: 797-805.

Gordon, R.H. and H.R. Varian (1988), Intergenerational risk sharing, Journal of Public Economics, 37: 185-202.

Grossman, G.M. and E. Helpman (1994), Protection for sale, American Economic Review, 84: 833-850.

Grossman, G.M. and E. Helpman (1996), Electoral competition and special interest politics, Review of Economic Studies, 63: 265-286.

Grossman, H.I. and S.J. Noh (1994), Proprietary public finance and economic welfare, Journal of Public Economics, 53: 187-204.

Guesnerie, R. (1975), Pareto optimality in non-convex economies, Econometrica, 43: 1-29.

Guesnerie, R. and K. Roberts (1984), Effective policy tools and quantity controls, Econometrica, 52: 59-86.

Guesnerie, R. and K. Roberts (1987), Minimum wage legislation as a second best policy, European Economic Review, 31: 490-498.

Guesnerie, R. and J. Seade (1982), Nonlinear pricing in a finite economy, Journal of Public Economics, 17: 157-180.

Harberger, A.C. (1962), The incidence of the corporation income tax, Journal of Political Economy, 70: 215-240.

Harberger, A.C. (1964), Taxation, resource allocation, and welfare, in: J. Due, ed., The Role of Direct and Indirect Taxes in the Federal Revenue System (Princeton University Press, Princeton) 25-70.

Harsanyi, J.C. (1955), Cardinal welfare, individualistic ethics and interpersonal comparisons of utility, Journal of Political Economy, 73: 309-321.

Haveman, R. (1994), Should generational accounts replace public budgets and deficits? Journal of Economic Perspectives, 8: 95-111.

Hemming, R. and M.J. Keen (1983), Single-crossing conditions in comparisons of tax progressivity, Journal of Public Economics, 20: 373-380.

Hettich, W. and S.L. Winer (1988), Economic and political foundations of tax structure, American Economic Review, 78: 701-712.

Hicks, U. (1946), The terminology of tax analysis, Economic Journal, 56: 38-50.

Hillier, B. and J.M. Malcomson (1984), Dynamic inconsistency, rational expectations, and optimal government policy, Econometrica, 52: 1437-1452.

Hindriks, J., M. Keen and A. Muthoo (1998), Corruption, extortion and evasion, Journal of Public Economics, forthcoming.

Hirshleifer, J. and J.G. Riley (1992), The Analytics of Uncertainty and Information (Cambridge University Press, Cambridge, U.K.).

Hochman, H.M. and J.D. Rodgers (1969), Pareto optimal redistribution, American Economic Review, 59: 542-557.

Hochman, H.M. and J.D. Rodgers (1973), Utility interdependence and income transfers through charity, in: K.E. Boulding, ed., Transfers in an Urbanized Economy (Wadsworth, Belmont, Calif.).

Hoff, K. (1994), The second theorem of the second best, Journal of Public Economics, 45: 223-242.

Hoff, K. and A.B. Lyon (1995), Non-leaky buckets: Optimal redistributive taxation and agency costs, Journal of Public Economics, 58: 365-390.

Hu, S.C. (1982), Social security, majority voting equilibrium and dynamic efficiency, International Economic Review, 23: 269-287.

Ihori, T. (1996), International public goods and contribution productivity differentials, Journal of Public Economics, 61: 139-154.

Inman, R.P. (1985), Markets, government and the new political economy, in: A. Auerbach and M.S. Feldstein, eds., Handbook of Public Economics: Vol. 2 (North-Holland, Amsterdam) 647-777.

Itaya, J.D., de Meza and G.D. Myles (1997), In praise of inequality: Public good provision and income distribution, Economics Letters, 57: 289-296.

Itsumi, Y. (1974), Distributional effects of linear income tax schedules, Review of Economic Studies, 41: 371-382.

Jakobsson, U. (1976), On the measurement of the degree of progression, Journal of Public Economics, 5: 161-168.

Janeba, E. and H. Raff (1997), Should the power to redistribute income be (de-)centralised? An example, International Tax and Public Finance, 4: 453-461.

Jha, R. (1998), Modern Public Economics (Routledge: London).

Johnson, W.R. (1977), Choice of Compulsory Insurance Schemes under Adverse Selection, Public Choice, 31: 23-35.

Kakwani, N. (1977), Application of Lorenz curves in economic analysis, Econometrica, 45: 719-727.

Kanbur, R. (1998), Income distribution and development, this volume, chapter 13.

Kanbur, R., M.J. Keen, and M. Tuomala (1994), Optimal non-linear income taxation for the alleviation of income poverty, European Economic Review, 38: 1613-1632.

Kay, J.A. (1980), The deadweight loss from a tax system, Journal of Public Economics, 13: 111-120.

Keen, M. J. (1996), The welfare economics of tax coordination in the European Community: A survey, in: M. Devereux ed., The Economics of Tax Policy (Oxford University Press, Oxford) 189-214.

Keen, M. (1997), Peculiar institutions: A British perspective on American tax policy, National Tax Journal, 50: 779-802.

Keen, M.J. (1998), The balance between specific and ad valorem taxation, Fiscal Studies, 19: 1-37.

Keen, M.J., H. Papapanagos and A.F. Shorrocks (1998), Tax reform and progressivity, Mimeo, University of Essex.

Koskela, E. and J. Vilmunen (1996), Tax progression is good for employment in popular models of trade union behaviour, Labor Economics, 3: 65-80.

Kotlikoff, L.J., (1984), Taxation and savings: A neoclassical perspective, Journal of Economic Literature, 22: 1576-1629.

Kotlikoff, L.J. (1992), Generational Accounting (The Free Press, New York).

Kotlikoff, L.J. (1987), Justifying public provision of social security, Journal of Policy Analysis and Management, 6: 674-689.

Kotlikoff, L.J. and L. Summers (1987), Tax incidence, in: A.J. Auerbach and M.S. Feldstein, eds., Handbook of Public Economics: Volume 2 (North-Holland, Amsterdam) 1043-1092.

Kotlikoff, L.J., T. Persson and L.E.O. Svenson (1988), Social contracts as assets: A possible solution to the time consistency problem, American Economic Review 78: 662-677.

Lambert, P.J. (1993), The Distribution and Redistribution of Income: a Mathematical Analysis (Edward Elgar, Aldershot).

Le Grand, J. (1982), The Strategy of Equality (Allen and Unwin, London).

Lecaillon, J., F. Paukert, C. Morrison and D. Germedis (1984), Income Distribution and Economic Development (International labor Organization, Geneva).

Lerner, A.P. (1944), The Economics of Control (Macmillan, New York).

Lillard, L.A. (1977), Inequality: Earnings vs. human wealth, American Economic Review, 67: 42-53.

Lindbeck, A. and J.W. Weibull (1987), Balanced-budget redistribution as the outcome of political competition, Public Choice, 52: 273-297.

Lindbeck, A. and J.W. Weibull (1993), A model of political equilibrium in a representative democracy, Journal of Public Economics, 51: 195-209.

Lindert, P. (1994), The rise of social spending 1880-1930, Explorations in Economic History, 31: 1-37.

Lindert, P. (1996), What limits social spending?, Explorations in Economic History, 33: 1-34.

Lindert, P. (1998), Three centuries of inequality in Britain and America, this volume.

Lockwood, B. and A. Manning (1993), Wage-setting and the tax system: Theory and evidence for the UK, Journal of Pubic Economics, 52: 1-29.

Lollivier, S. and J.-C. Rochet (1983), Bunching and second-order conditions: A note on optimal tax theory, Journal of Economic Theory, 31: 392-400.

Magee, S.P., W.A. Brock and L. Young (1989), Black hole Tariffs and Endogenous Policy Theory (Cambridge University Press, Cambridge, UK).

Marceau, N. and R.W. Boadway (1994), Minimum wage legislation and unemployment insurance as instruments for redistribution, Scandinavian Journal of Economics, 96: 67-81.

Marglin, S.A. (1963), The social rate of discount and the optimal rate of investment, Quarterly Journal of Economics, 77: 95-112.

Marhuenda, F. and I. Ortuño-Ortín (1997), Tax enforcement problems, Scandinavian Journal of Economics, 99: 61-72.

Mas-Colell, A., M.D. Whinston and J.R. Green (1995), Microeconomic Theory (Oxford University Press, Oxford).

May, K.O. (1952), A set of independent, necessary and sufficient conditions for simple majority decision, Econometrica, 20: 680-684.

Mayshar, J. and S. Yitzhaki (1995), Dalton-improving indirect tax reform, American Economic Review, 85: 793-807.

McGuire, M. and M. Olsen Jr. (1996), The economics of autocracy and majority rule: The invisible hand and the use of force, Journal of Economic Literature, XXXIV: 72-96.

Meltzer, A.H. and S.F. Richard (1981), A rational theory of the size of government, Journal of Political Economy, 89: 914-927.

Meltzer, A.H. and S.F. Richard (1983), Tests of a rational theory of the size of government, Public Choice, 41: 403-418.

Mieszkowski, P.M. (1969), Tax incidence theory: the effects of taxes on the distribution of income, Journal of Economic Literature, 7: 1103-1124.

Mintz, J.A. (1994), Is there a future for capital income taxation?, Canadian Tax Journal, 42: 1469-1503.

Mirrlees, J.A. (1971), An exploration in the theory of optimum income taxation, Review of Economic Studies, 38: 175-208.

Mirrlees, J.A. (1974), Notes on welfare economics, information, and uncertainty, in: M. Balch, D. McFadden and S. Wu, eds., Essays on Equilibrium Behavior under Uncertainty (North-Holland, Amsterdam) 243-258.

Mirrlees, J.A. (1990), Taxing uncertain incomes, Oxford Economic Papers, 42: 34-45.

Mookherjee, D. and I. Png (1989), Optimal auditing, insurance, and redistribution, Quarterly Journal of Economics, 104: 399-415.

Mueller, D. (1989), Public Choice II (Cambridge University Press; Cambridge, UK).

Munro, A. (1992), Self-selection and optimal in-kind transfers, Economic Journal, 102: 1184-1196.

Murphy, K.M., A. Schleifer and R.W. Vishny (1989), Income distribution, market size, and industrialization, Quarterly Journal of Economics 104: 537-64.

Musgrave, R.A. (1959), The Theory of Public Finance (McGraw-Hill, New York).

Musgrave, R.A., K.E. Case and H.B. Leonard (1974), The distribution of fiscal burdens and benefits, Public Finance Quarterly, 2: 259-311.

Myles, G.D. (1988), Some implications of quality differentials for optimal taxation, Economic Journal, 98: Conference Issue, 148-160.

Myles, G. D. (1995), Public Economics (Cambridge University Press, Cambridge, UK).

Nava, M., F. Schroyen and M. Marchand (1996), Optimal fiscal and public expenditure policy in a two-class economy, Journal of Public Economics, 61: 119-137.

Neal, D. and S. Rosen (1998), Theories of the distribution of earnings, this volume, chapter 7.

Nichols, A.L. and Zeckhauser, R.J. (1982), Targeting transfers through restrictions on recipients, American Economic Review, 72: 372-377 .

Nicholson, J.L. and A.J.C. Britton (1976), The redistribution of income, in: A.B.Atkinson, ed., The Personal Distribution of Income (Allen and Unwin, London) 313-334.

Nozick, R. (1974), Anarchy, State and Utopia (Basic Books, New York).

Oates, W.E. (1972), Fiscal Federalism (Harcourt Brace Jovanovich, New York).

Okun, A.M. (1975), Equality and Efficiency (The Brookings Institution, Washington).

Olsen, M. Jr. (1965), The Logic of Collective Action (Harvard University Press, Cambridge, MA).

Olsen, M. Jr. (1982), The Rise and Decline of Nations: Economic Growth, Stagflation, and Social Rigidities (Yale University Press, New Haven, Conn.).

Osborne, M.J. and Slivinski, A. (1996), A model of political competition with citizen-candidates, Quarterly Journal of Economics, 111: 65-96.

Oswald, A.J. (1983), Altruism, jealousy and the theory of optimal non-linear taxation, Journal of Public Economics, 20: 77-87.

Parsons, D.O. (1996), Imperfect 'Tagging' in Social Insurance Programs, Journal of Public Economics, 62: 183-207.

Pattanaik, P. and A.K. Sen (1969), Necessary and sufficient conditions for rational choice under majority decision, Journal of Economic Theory, 1: 178-202.

Pauly, M.V. (1973), Income redistribution as a local public good, Journal of Public Economics, 2: 35-58.

Pechman, J.A. and B.A. Okner (1974), Who Bears the Tax burden? (The Brookings Institution, Washington).

Peltzman, S. (1980), The growth of government, Journal of Law and Economics, 23: 209-287.

Persson, M. (1995), Why are taxes so high in egalitarian countries?, Scandinavian Journal of Economics, 97: 569-580.

Persson, T. and G. Tabellini (1994a), Is inequality harmful for growth?, American Economic Review, 84: 600-621.

Persson, T. and G. Tabellini (1994b), Representative democracy and capital taxation, Journal of Public Economics, 55: 52-70.

Pestieau, P. (1998), The political economy of social security reform, Mimeo, University of Liège.

Piggott, J.R. and J. Whalley (1985), U.K. Tax Policy and Applied General Equilibrium Analysis (Cambridge University Press, Cambridge, UK).

Pigou, A.C. (1947), A Study in Public Finance: 3rd Edition (Macmillan, London).

Piketty, T. (1998), Theories of persistent inequality and intergenerational mobility, this volume, chapter 8.

Piketty, T. (1997), The dynamics of the wealth distribution and the interest rate with credit rationing, Review of Economic Studies, 64: 173-189.

Plott, C.R. (1967), A notion of equilibrium and its possibility under majority rule, American Economic Review, 57: 787-806.

Pomp, R.D. (1989), The state, the individual and the taxation of economic migration, in: J.N. Bhagwati and J.D. Wilson, eds., Income Taxation and International Mobility (MIT Press, Cambridge).

Poterba, J.M. (1989), Lifetime incidence and the distributional burden of excise taxes, American Economic Review, 79: 325-330.

Prest, A.R. (1955), The statistical calculation of tax burdens, Economica, 22: 234-245.

Pyne, D. (1997), Microfoundations of influencing public opinion: Lobbying and voting for trade policies, Mimeo, University of York.

Rae, D.W. (1969), Decision-rules and individual values in constitutional choice, American Political Science Review, 63: 40-56.

Ravallion, M. (1984), How much is a transfer payment worth to a rural worker?, Oxford Economic Papers, 36: 478-489.

Ravallion, M. (1987), Towards a theory of famine relief policy, Journal of Public Economics, 33: 21-39.

Rawls, J. (1971), A Theory of Justice (Harvard University Press, Cambridge, MA).

Reinganum, J.F. and L.L. Wilde (1985), Income tax compliance in a principal-agent framework, Journal of Public Economics, 26: 1-18.

Renstrom, T.I. (1997), Tax evasion as a disciplinary mechanism for fiscal policy, paper presented at the 53rd Congress of the International Institute of Public Finance, Kyoto Japan.

Roberts, K.W.S. (1977), Voting over income tax schedules, Journal of Public Economics, 8: 329-340.

Roberts, K.W.S. (1984), The theoretical limits to redistribution, Review of Economic Studies, LI: 177-195.

Rodrik, D. (1986), Tariffs, subsidies, and welfare with endogenous policy, Journal of International Economics, 21: 285-299.

Rodrik, D. (1995), Political economy of trade policy, in: G. Grossman and K. Rogoff, eds., Handbook of International Economics (North-Holland: Amsterdam).

Rodrik, D. (1998), Why do more open economies have bigger governments?, Journal of Political Economy, 106: 997-1032.

Roell, A. (1996), Voting over non-linear income tax schedules, Mimeo, ECARE.

Roemer, J. (1996), Egalitarian Perspectives: Essays in Philosophical Economics (Cambridge University Press, Cambridge, UK).

Roemer, J.E. (1998), Why the poor do not expropriate the rich: An old argument in new garb, Journal of Public Economics, 70: 399-424.

Roemer, J.E. (1999), The democratic political economy of progressive income taxation, Econometrica, 67: 1-19.

Rogers, C.A. (1987), Expenditure taxes, income taxes and time-inconsistency, Journal of Public Economics, 32: 215-230.

Romer, T. (1975), Individual welfare, majority voting and the properties of a linear income tax, Journal of Public Economics, 4: 163-185.

Rose-Ackerman, S. (1996), Altruism, nonprofits, and economic theory, Journal of Economic Literature 34: 701-728.

Rothschild, M. and J.E Stiglitz (1976), Equilibrium in competitive insurance markets, Quarterly Journal of Economics, 90: 629-650.

Sadka, E. (1976), On progressive income taxation, American Economic Review, 66: 931-935.

Sah, R.K. (1983), How much redistribution is possible through commodity taxes?, Journal of Public Economics, 20: 89-101.

Sala-i-Martin, X. (1997), Transfers, social safety nets and economic growth, IMF Staff Papers, 44: 81-102.

Seade, J.K. (1977), On the shape of optimal income tax schedules, Journal of Public Economics, 7: 203-235.

Seade, J.K. (1985), Profitable cost increases and the shifting of taxation: Equilibrium responses of markets in oligopoly, University of Warwick Discussion Paper No. 260.

Scotchmer, S. and J. Slemrod (1989), Randomness in tax enforcement, Journal of Public Economics, 38: 17-32.

Sen, A.K. (1966), A possibility theorem on majority decisions, Econometrica, 34: 491-499.

Sen, A.K. (1967), Isolation, assurance and the social rate of discount, Quarterly Journal of Economics, 81: 112-124.

Sen, A.K. (1970a), Collective Choice and Social Welfare (Oliver and Boyd, Edinburgh).

Sen, A.K. (1970b), The impossibility of a Paretian liberal, Journal of Political Economy, 78: 152-157.

Sen, A.K. (1973), On Economic Inequality (Clarendon Press, Oxford).

Sen, A.K. (1977), Social choice theory: A re-examination, Econometrica, 45: 53-89.

Sen, A.K. (1985), Commodities and Capabilities (North-Holland, Amsterdam).

Sen, A.K. (1992), Inequality Reexamined (Harvard University Press, Cambridge, Mass.).

Sen, A.K. (1995), The political economy of targeting, in: D. van de Walle and K. Nead, eds., Public Spending and the Poor (Johns Hopkins University Press; Baltimore) 11-24.

Shah, A. and J. Whalley (1991), The redistributive impact of taxation in developing countries, in: J. Khalilzadeh and A. Shah, eds., Tax Policy in Developing Countries (World Bank, Washington DC).

Sheshinski, E. (1972), The optimal linear income tax, Review of Economic Studies, 39: 297-302.

Sheshinski, E. (1989), Note on the shape of the optimum income tax schedule, Journal of Public Economics, 40: 201-215.

Shoven, J.B. and J. Whalley (1972), A general equilibrium calculation of the effects of differential taxation of income from capital in the U.S., Journal of Public Economics, 1: 281-321.

Shoven, J.B. and J. Whalley (1984), Applied general-equilibrium models of taxation and international trade, Journal of Economic Literature, 22: 1007-1051.

Sinn, H-W. (1989), Tax harmonization and tax competition in Europe, European Economic Review, 34: 489-504.

Sinn, H-W. (1995), A theory of the welfare state, Scandinavian Journal of Economics, 97: 495-526.

Sinn, H-W. (1996), Social insurance, incentives and risk-taking, International Tax and Public Finance, 3: 259-280.

Sjoblom, K. (1985), Voting for social security, Public Choice, 45: 225-240.

Slemrod, J. (1990), Optimal taxation and optimal tax systems, Journal of Economic Perspectives, 4: 157-178.

Slemrod, J., S. Yitzhaki, J. Mayshar, and M. Lundholm (1996), The optimal two-bracket linear income tax, Journal of Public Economics, 53: 1996, 269-290.

Sørensen, P.B. (1998), Public finance solutions to unemployment problems, paper presented at the 54th Congress of the International Institute of Public Finance, Cordoba, Argentina.

Starrett, D. (1972), On golden rules, the 'biological rate of interest', and competitive efficiency, Journal of Political Economy, 80: 276-91.

Stern, N. (1976), On the specification of models of optimum income taxation, Journal of Public Economics, 6: 123-162.

Stern, N. (1982), Optimum taxation with errors in administration, Journal of Public Economics, 17: 181-211.

Stern, N. (1987), The effects of taxation, price control and government contracts in oligopoly and monopolistic competition, Journal of Public Economics, 32: 133-158.

Stigler, G.J. (1970), Director's law of public income redistribution, Journal of Law and Economics, 13: 1-10.

Stigler, G. (1982), Economists and public policy, Regulation, 6: 13-17.

Stiglitz, J.E. (1982), Self-selection and pareto efficient taxation, Journal of Public Economics, 17: 213-240.

Stiglitz, J.E. (1987), Pareto efficient and optimal taxation and the new new welfare economics, in: A.J. Auerbach and M. Feldstein, eds., Handbook of Public Economics, Volume II (North-Holland, Amsterdam) 991-1042.

Stiglitz, J.E. (1998), Taxation, public policy, and the dynamics of unemployment, paper presented at the 54th Congress of the International Institute of Public Finance, Cordoba, Argentina.

Summers, L.H. (1981), Capital taxation and accumulation in a life cycle growth model, American Economic Review, 71: 533-544.

Summers, L.H. (1989), Some simple economics of mandated benefits, American Economic Review, Papers and Proceedings, 79: 177-183.

Tabellini, G. (1991), The politics of intergenerational redistribution, Journal of Political Economy, 99: 335-357.

Thurow, L.C. (1971), The income distribution as a pure public good, Quarterly Journal of Economics, 85: 416-424.

Townley, P.G. (1981), Public choice and the social insurance paradox: A note, Canadian Journal of Economics, 14: 712-717.

Tullock, G. (1967), The welfare costs of tariffs, monopolies and theft, Western Economic Journal, 5: 224-232.

Tullock, G. (1983), Further tests of a rational theory of the size of government, Public Choice, 41: 419-421.

Tullock, G. (1997), Economics of Income Redistribution (Kluwer Academic, Boston).

Usher, D. (1977), The welfare economics of the socialization of commodities, Journal of Public Economics, 8: 151-168.

Usher, D. (1986), Tax evasion and the marginal cost of public funds, Economic Inquiry, 24: 563-586.

Usher, D. and M. Engineer (1987), The distribution of income in a despotic society, Public Choice, 54: 261-276.

van de Walle, D. and K. Nead (1995), Public Spending and the Poor: Theory and Evidence (Johns Hopkins University Press, Baltimore).

Van Graaff, J. de (1957), Theoretical Welfare Economics (Cambridge University Press, Cambridge, UK).

Varian, H.R. (1980), Redistributive taxation as social insurance, Journal of Public Economics, 14: 49-68.

Verbon, H.A.A. (1987), The rise and evolution of public pension systems, Public Choice, 52: 75-100.

Vermaeten, A., W.I. Gillespie and F. Vermaeten (1995), Who paid the taxes in Canada, 1951-1988?, Canadian Public Policy, 21: 317-343.

Vickrey, W. (1945), Measuring marginal utility by attitudes towards risk, Econometrica, 13: 215-236.

Vigneault, M. (1996), Commitment and the time structure of foreign direct investment, International Tax and Public Finance, 3: 479-494.

Warr, P.G. (1983), The private provision of a public good is independent of the distribution of income, Economics Letters, 13: 207-211.

Weitzman, M. (1977), Is the price system or rationing more effective in getting a commodity to those who need it most?, Bell Journal of Economics, 8: 517-524.

Wen, J.-F. (1997), Tax holidays and the international capital market, International Tax and Public Finance, 4: 129-148.

Weymark, J.A. (1986), Bunching properties of optimal nonlinear income taxes, Social Choice and Welfare, 3: 213-232.

Whalley, J. (1984), Regression or progression: the taxing question of tax incidence?, Canadian Journal of Economics, 17: 654-682.

Wildasin, D.E. (1991), Income distribution in a common labor market, American Economic Review, 81: 757-774.

Wildasin, D.E. (1998), Factor mobility and redistributive policy: Local and international perspectives, in: P.B. Sorensen, ed., Public Finance in a Changing World (MacMillan Press, London).

Wilson, C. (1980), The nature of equilibrium in markets with adverse selection, Bell Journal of Economics, 11: 108-130.

Wilson, J.D. (1990), Are efficiency improvements in government transfer policies self-defeating in political equilibrium?, Economics and Politics, 2: 241-258.

Wilson, L.S. and M.L. Katz (1983), The socialization of commodities, Journal of Public Economics, 20: 347-356.

Wright, R. (1996), Taxes, redistribution, and growth, Journal of Public Economics, 62: 327-338.