16

INFORMATION AND INCENTIVES: THE ECONOMICS OF CARROTS AND STICKS

Nobel Lecture, December 9, 1996

by

JAMES A. MIRRLEES

Faculty of Economics and Politics, University of Cambridge, England

THE INVISIBLE HAND

In a lecture that will deal chiefly with ignorance, it may seem natural to begin with Adam Smith's most famous contribution to economics, his vision of independent selfish beings who by living and working together in the economic system somehow do what is best for one another. First, in *The Theory of Moral Sentiments*, he said

> The rich only select from the heap what is most precious and agreeable. They consume little more than the poor, and in spite of their natural selfishness and rapacity ...they divide with the poor the produce of all their improvements. They are led by an invisible hand to make nearly the same distribution of the necessaries of life which would have been made, had the earth been divided into equal portions among all its inhabitants. (IV.i.10).

This is far from the later conception of an economic equilibrium that is "optimal" in Pareto's sense. Indeed as quoted, Smith's early claim is not very plausible. It does set the major themes: the working of the economy as a system, and the good or otherwise, for everyone, that can flow from it. Later in the *Wealth of Nations*, he argued correctly that individual profit-maximization implies maximization of what we would call national income, and goes on to say that,

> ...by directing that industry in such a manner as its produce may be of the greatest value, [every individual] intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention. (IV, Chapter II).

This says nothing about possible advantage to the poor, indeed nothing about the distribution of gains at all.

As taught to generations of economists, there are two parts to the doctrine of the invisible hand. The first is that an economic equilibrium is Pareto-optimal: it is an allocation of commodities and activities to people with the property that no other allocation could make everyone better off. It is a good allocation in rather a weak sense, but better for everyone than a lot of other possible allocations. The economic equilibrium has to be perfectly competitive. The second part says that any Pareto-optimal allocation can be an economic equilibrium. For that to be, the initial distribution of assets among

people has to be set right. It may be required that the earth is indeed divided into equal portions among its inhabitants for the desired allocation to emerge. This second proposition, at least in its standard form, makes assumptions about the nature of technological possibilities such that a perfectly competitive equilibrium can occur. Essentially economies of scale have to be excluded, or production levels in such industries determined in some other way, for example by some kind of planning. That is an interesting issue, but will be ignored here.[1]

These propositions were the essential content of welfare economics as I learned it in the fifties. Ian Little [1950] and particularly Jan Graaff [1951] brought out the many serious difficulties in the theory, particularly if it were to be taken as a basis for economic advice and policy and ideology. The theory underlay much of what economists thought they could tell the world and its rulers. It was the basis for free trade arguments, for urging the control of monopoly, for methods of cost-benefit analysis, and the justification of marginal-cost pricing by publicly-owned firms. It was also used to support the extension of free markets and private ownership of property, and to recommend the use of price systems even in planned economies.

Yet the defects of the theory seemed serious. Many economic transactions take place between individual agents or firms, with significant monopoly on at least one side of the market. Taking a later view, that seems to be because of search costs and switching costs and uncertainty about the fulfilment of contracts in the future (as in credit markets). These have no place in the competitive model of the economy. One might possibly be able to claim that these deviations from the assumptions are small, within the margins of error that economics can aspire too; though for myself I do not think they are.

The other major defect is the need to have a particular distribution of assets to people before one can claim that the resulting equilibrium is good. That requirement, when properly understood, was plainly impossible to fulfil.

What is the nature of this difficulty? It had, I think, become quite clear by the fifties. It was clear in William Vickrey's writings, for example[2]. If we are to have a good equilibrium, we shall have to imagine a good government that does what is needed, namely to create a distribution of assets to people such that the desired allocation is indeed an equilibrium. It can be done. In one exceedingly simple model, which perhaps many economists had in mind, people are all the same, and each person obtains utility from a single consumption good, of which a fixed amount is available. Then it is easy for the government to do it. Assuming diminishing marginal utility of consumption, and comparability of individual utilities, an equal distribution of assets is what we require. No information, other than a census, is required for that. People might have some doubts about the measurability, perhaps even about the meaningfulness, of utility; but at least in a rough and ready way, there was a

[1] There are ways of extending the proposition, discussed in Mirrlees[1995], and work referred to therein.

[2] The problem is discussed, somewhat elliptically, in Vickrey[1945], and that paper puts forward essentially the model of redistributive taxation discussed later in this lecture.

strong case for thinking that transfer from richer to poorer was an improvement. Carrying that to the logical extreme, the riches of the earth should be equally distributed.

It was not a popular policy, in part for good reasons. Obviously if a perfectly equalizing policy were carried out, the ordinary incentive to work would be eliminated. "From each according to his abilities, to each according to his needs." (Karl Marx, *Criticism of the Gotha Programme*) is not thought to be feasible, even if desirable. Nothing in the simple model allowed for that.

TAXATION

What exactly was the problem? In general, and in reality, the redistribution of assets required by the first welfare theorem needed information the government could not obtain. This ideal government had to know what wealth people already had, and what they were capable of doing, before it could work out how much to give and take. If people knew it needed that information, they could in one way or another dissemble, and would if it benefitted them to do so. The information requirements of the second welfare theorem cannot be fulfilled. Transfers to or from a person that depend on the characteristics of that person, not his behaviour, are known as lump-sum transfers. Desirable lump-sum transfers are, in effect, impossible, because they require information that is not available. The attempt to implement them would be expected to destroy the value of the information on which they would have to be based.

Following that line of thought, in the mid-sixties, Peter Diamond and I were convinced that one should think about economic welfare and economic policy in the context of public finance. At first we studied a general economic model in which the government was not able to use lump-sum transfers at all[3]. Clearly that was going too far, and we went on to allow that the government could use uniform lump-sum taxes or, more plausibly, subsidies. Otherwise it had to use taxes, taxes that from the point of view of the pure welfare theorems are regarded as distorting. This was a model in which all consumers and producers were price takers, but they did not necessarily face the same prices, because tax rates could make the two parties to a transaction face different effective prices. It was a conventional model, in having competitive behaviour of private agents; but it was a distorted economy, and the distortion could be optimal. That was what we wanted to study.

There was a significant earlier literature on distorting taxes. The theory had been started by Frank Ramsey [1927], at the request of A. C. Pigou. He considered an economy of identical consumers and assumed that revenue had to be raised entirely through commodity taxes. M. Boiteux [1956] in France had later, but independently, developed the isomorphic theory of pricing by a public utility with a zero-profit constraint, in an economy where, pa-

[3] The theory we developed was eventually published in two papers, Diamond and Mirrlees[1971].

radoxically, perfect lump-sum transfers were supposed to be happening. And Serge-Christophe Kolm [1971] also developed a systematic general theory.

What is interesting for the story I have to tell now is the form of the theory we developed, rather than its content and implications. It was the theory of a government whose actions were a function only of what it could observe. No observations about the nature of individual consumers were assumed to be available, only observations of their behaviour. The leading example of that is a commodity tax, which provides revenue proportional to the quantity of the commodity consumers choose to buy. Policy choice takes account of consumer responses to tax changes, namely changes in amounts purchased. A uniform lump-sum subsidy requires no information at all, and is therefore allowed. All these policy instruments are, as we say now, incentive-compatible: the government takes full account of people's self-interested response to the tax system.

There is one respect in which information about consumer characteristics is needed, the distribution of these characteristics within the population. The government somehow knows this distribution. In principle it could obtain it by asking people, or putting them through various tests. Since the information is used only in aggregate form, an individual has nothing to lose (or gain) by telling the government the truth. Only if individual tax liability were affected by these revelations would incentive-compatibility be violated. We supposed that the government would obtain and use an econometric model of consumers, in which the distribution of consumer types–their intensity of taste for different commodities, their labour supply characteristics–would be estimated from the behaviour of a sample of consumers. Calculations of optimal taxes and of desirable directions of tax change have been done on in that way. One should allow explicitly for uncertainties about the distribution of characteristics in the population, with the budget-balancing issues it raises, but we did not trouble about what is a relatively minor complication.

This then was a model with asymmetric information, where, at the time when government policy is determined, individual consumers know more about their tastes and abilities than the government does. Since all kinds of policy parameters could be mentioned in the model, it was quite general. That generality was not fully exploited.

INCOME TAXATION

It is interesting to be more specific, and study a particular economic model that has some of the most salient features of real economies. One such model pictures the economy as timeless, with people enjoying a single consumption good, and supplying labour. As mentioned above, William Vickrey had stated the optimum tax problem for such a model, though he was not able to solve it. In such a model, redistributive taxation can be described simply as an income tax, which has the effect of determining each individual's consumption as a function of that individual's labour income.

The point about income taxation is that tax can be a highly complicated

function of income: tax need not be proportional to income, and in the real world seldom is. In the real economy, income actually consists of several elements. It is often easy to distinguish between labour income and income from capital, conceptually at least. (In practice, particularly with the self-employed, the distinction might be hard to enforce; and net income from capital, including housing, can itself be hard to measure.) In what follows, I shall assume we are dealing with labour income. Labour income can be taxed nonlinearly. So can a number of other commodities, such as telephone calls, electricity and the like. Fully nonlinear taxation is, like linear taxation, incentive-compatible: the calculation of tax is still based on the publicly observable behaviour of the consumer. The presumption is that nonlinearity will be advantageous, because more general than linear taxation.

In practical terms, one could not tell in advance how advantageous nonlinearities might be. Most countries had and have large differences in marginal labour-income tax rates, and in most cases the income-tax law specifies marginal tax rates that increase with income. There are also other elements in the social insurance and tax arrangements of countries that are very like a form of income taxation, for example unemployment insurance, or low-income support arrangements. In many cases, these have the effect of creating quite high marginal tax rates on low incomes: benefits are reduced as income increases, sometimes almost one-for-one. The typical real tax system therefore has high marginal tax rates for low incomes and for high incomes, and low marginal tax rates in the intermediate range. It is tempting to think of most taxes, other than taxes on capital income, as amounting to a tax on labour income: that would be right if different consumer goods were taxed at the same rate. At any rate one can approximate real tax systems quite closely in these terms.

The next step in thinking about an optimum income-tax system was, paradoxically, to move back from thinking explicitly about tax rates, and instead to think about allocations of real commodities and labour. It proved to be advantageous to go back to thinking about optimality as in the general welfare economics from which all this work had begun, but now to have a new kind of constraint, incentive-compatibility, in addition to the constraint that the allocation should be feasible. The idea of incentive-compatible transfers by the government had been captured by thinking in terms of tax rates. But it is really a more fundamental idea. The question to ask was: what allocations are possible if policy has to be incentive-compatible? It seemed best to think about this in a special model, but in fact the answer turned out to be quite simple in a fairly general model.

The special model was one that had long seemed natural to me, in which individual consumers can choose how much labour to supply. Each consumer's situation is described by two variables, consumption and labour supply. Each consumer's type is defined by a single parameter, productivity, or, equivalently, that person's wage rate. There is a given distribution of wage rates in the economy, known to the government. These are real, the actual productivity of the different individuals. The government can observe the total pro-

duct of each individual, that is the product of the wage rate and the amount worked, but is unable to observe either of these alone. That observability assumption is a bit extreme, and I shall come back to it. But there are certainly severe limits on what the government can observe, and this particular assumption corresponds to what tax systems almost invariably do: they relate only to total income, not to wage rates.

The government was also supposed to have an aim, a measure of welfare that it wishes to maximize, a sum of individual utilities, consistent with individual preferences for consumption and work. That does not matter for the first step in the analysis, which is to find a way of describing the real allocations that are possible for the government, that is to say the real distributions of consumption and work in the population that are incentive-compatible. These are the allocations that are possible with some labour-income tax system, but I wanted to describe what allocations were possible without reference to taxes, and that was the essential step to having a computable model of general taxes.

I have said the answer was simple. It is shown in Figure 1. For each consumer call the product of wage and labour, income. Incentive-compatibility required that each consumer would choose from a set of available consumption income pairs. That set is defined by the allocation of consumption and income among consumers. A curve, labelled BB′ in the diagram, describes that allocation, showing consumption at different income levels. Each consumer chooses from that curve: each has an indifference curve tangential to the allocation curve, such as II′ and JJ′ in Figure 1. To be more precise, what I have called a curve might, technically, not be: it could well have corners. Still, it followed from this simple argument that the curve must be a lower envelope to a collection of individual indifference curves, one for each type of consumer. As a consequence, utility increased as the wage increased, at a rate equal to the derivative of utility with respect to the wage, holding consumption and income constant. And also income was an increasing function of the consumer's wage rate. These two facts together fully characterized the set of incentive-compatible consumption/income allocations.

One key assumption was needed to justify that conclusion. It had to be assumed that people with higher wage rates always found it easier to produce more income (by working) than those with smaller. That is more restrictive than it sounds, and is much more than a definition of increasing wage, but seems an entirely reasonable and plausible assumption. In the Figure, the assumption means that different people's indifference curves cross one another once only. The condition is known as the single-crossing property (or sometimes the Spence-Mirrlees condition[4]). With that assumption, one had a full characterization of incentive-compatible allocations.

Furthermore, and crucially, the original optimal-income-tax problem could now be converted into something very like a standard control-theory problem, with utility as the state variable, income the control variable. The enve-

[4] Michael Spence [1973] used the condition for models of markets with asymmetric information.

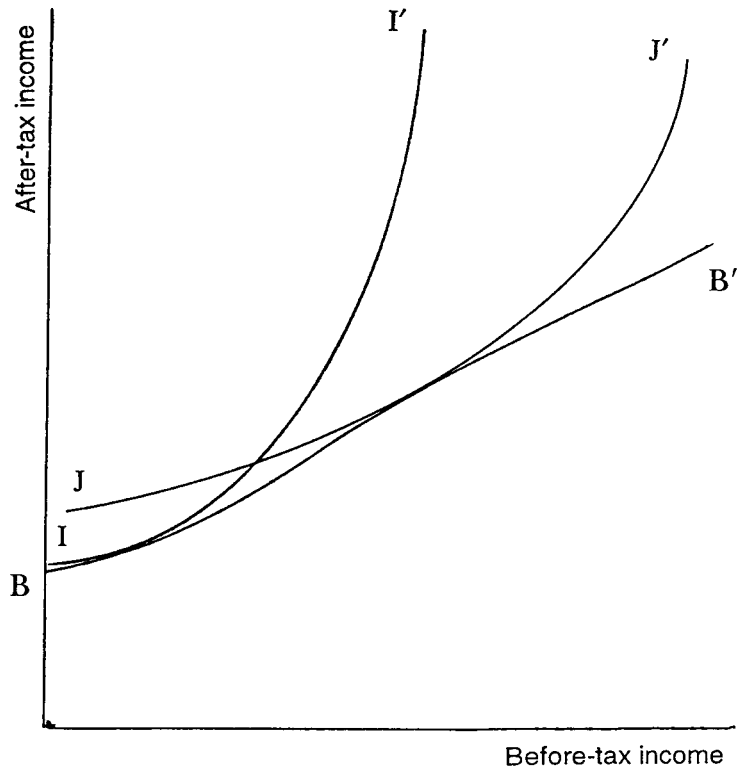


Figure 1. Incentive-compatibility.

lope condition just described was essentially equivalent to a statement that the rate at which utility increased in the population, with respect to the wage rate, was equal to the partial derivative of the individual's utility with respect to the wage rate, just a known function of that individual's consumption, income, and wage. To tell a little more of the truth, one has to generalize all this somewhat, for utility may not always be a smooth function of the wage. Consequently the full mathematical justification of all this is quite complex. Computation of results was not trivial either, because in fact one had to check whether there were ranges of consumers all of whom would make the same income; but it could be done[5].

An exciting feature of this analysis, and one that came as a complete surprise, was its validity. That puts the issue too starkly. Let us rather say that the use of fully nonlinear taxation does solve one problem that has troubled tax theorists. In the optimal commodity tax analysis that Peter Diamond and I had done, we had obtained first-order conditions for optimal taxes, and explored various interpretations and implications; but these conditions were necessary conditions for optimality, not sufficient conditions. In any particular model, computation of optimal taxes should require much more than solving

[5] All of this is set out in Mirrlees[1971], while the full mathematical justification (in a Nuffield College, Oxford discussion paper) appeared in print only in Mirrlees[1986].

the first-order conditions, unless by good fortune there were a unique solution. That problem need not afflict simple welfare economics with perfect lump-sum transfers. In the income-tax problem, relatively simple conditions, easily checked for the particular model I was using, implied that the solution of the equations did give an optimum: the conditions were sufficient as well as necessary. When the computations were done, one knew one had the right answer, not just an answer that might be right.

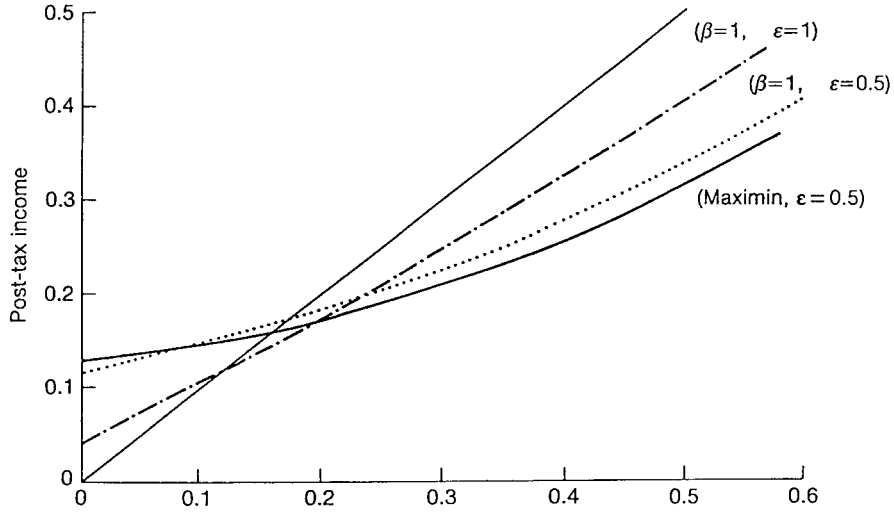
Solution of the model in any particular case shows how consumption should be related to income. From that one can talk about income tax rates, interpreting the difference between income and consumption as tax. Remember that in the model, the allocation would be achieved with just that tax. The income tax in the model corresponds to the sum of the real-world labour-incomes taxes and taxes on consumption, such as value-added tax.

Computation of the model was done for particular cases. There are three key assumptions: the distribution of wages, the nature of individual preferences between consumption and work, and the extent to which it is supposed desirable to transfer from the better off to the less well off, i.e. the way that the welfare function incorporates individual preferences. In the 1971 paper, the simplest reasonable assumption was made about consumption/work preferences, namely unit elasticity of substitution between consumption and leisure. At least for male workers, work since then suggests that elasticity is considerably too high. Later work[6] shows that marginal rates of tax should as a consequence be greater than they were in these first calculations. The distribution of labour incomes is not all that easy to observe. In any case intertemporal aspects are important in the real world and completely absent in the model. Both log-normal distributions and distributions with Pareto tails were tried, and gave distinctly different results, particularly at upper incomes. Different welfare specifications had much more effect at low incomes than high, and there are good theoretical grounds for that. A required level of public expenditure (on defence, police, etc., not welfare spending, which is part of the optimal tax system) was also postulated.

The kinds of results obtained are illustrated in Figures 2 and 3, giving calculations by Matti Tuomala. The two parameters, β and ε, describe the degree of egalitarianism assumed, and the elasticity of substitution between consumption and labour-supply. $\varepsilon = 1$ corresponds to the cases calculated in the 1971 paper, $\varepsilon = 0.5$ is probably a more realistic value.

Several things were striking about the results. In many cases, marginal tax rates were highest in the middle of the range of incomes, and fell towards higher incomes and lower. This was the opposite of actual tax systems. It was a feature that at that time seemed quite robust, though later computations and results suggest that marginal tax rates can be quite high at the lowest incomes. (With a very highly egalitarian form of welfare judgment, the marginal tax rate appears to fall all the way up the income range.) Another result, not numerically very striking, but on consideration important, was that it was

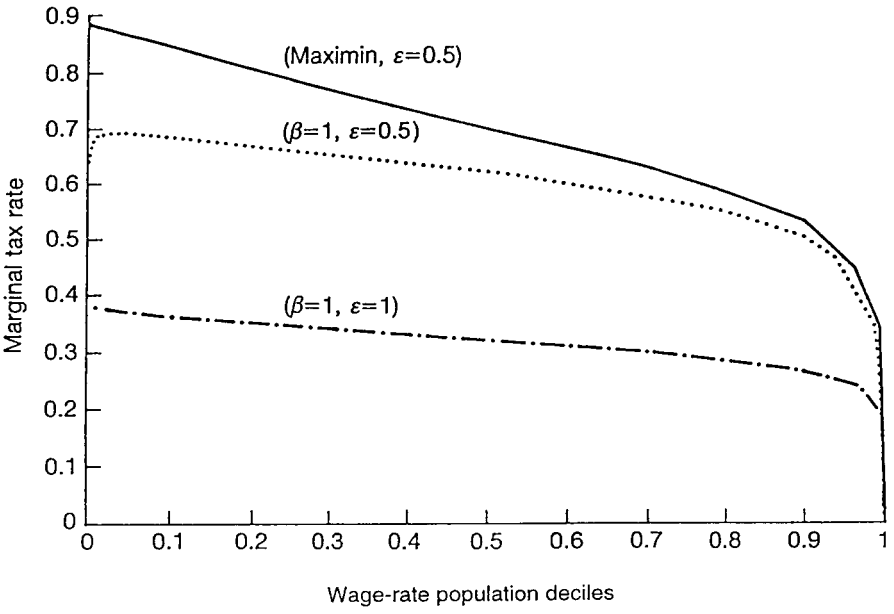
[6] Tuomala[1990].



β = 1 means welfare is the sum of utilities
ε is the consumer elasticity of substitution between consumption and labour

Source: M. Tuomala, *Journal of Public Economics* (1984).

Figure 2. Optimal Income-tax Schedules.



Source: M. Tuomala, *Journal of Public Economics* (1984).

Figure 3. Marginal Tax Rates.

optimal to have a positive amount of unemployment. People who chose to earn no income at all are paid a subsidy, since we did not wish them to starve, and those with very low productivity therefore found pay insufficient to justify working.

One can get some insight into the problem and the results by thinking about an extreme case where inequality of incomes is rather low. The limiting case is where everyone has the same income. Then there is no incentive-compatibility problem. It is optimal to raise the funds required for pure public expenditure by a lump-sum tax. The marginal tax rate is zero. Now let there be a little inequality, with wages ranging all the way from zero to some high level, but a small variance. For people in the middle of the wage range (who are most of the population) taxes should be very similar to what they were in the equal-wages case, i.e. a low marginal tax rate, and an average tax rate big enough to pay for public expenditure. But that cannot apply to people on the lowest incomes, because they could not have paid that lump-sum tax–it would have meant negative consumption. The consumption/income schedule must therefore be curved at the lower end, always remaining below the indifference curve of an average-wage earner. One cannot see from this argument quite how low the gradient of the consumption/income schedule should be at the lower end, but it will surely be a lot lower than at the average income, which is to say that the marginal tax rate will be much higher at the lower end than in the middle.

As the wage distribution gets more unequal, the optimal consumption/income schedule changes in shape in quite a complicated way. Another possible reason for high marginal tax rates in the lower range can come into play. If people in fact like to do some work, it may not be so important to provide labour incentives to people with low wage rates. That can mean tax rates at or close to 100 per cent. can be optimal at the bottom of the income range. In higher ranges, recent work of Peter Diamond, as yet unpublished, shows that the inverted U-shape of the consumption/income schedule is quite common with marginal tax rates rising at the mode of the distribution, but eventually falling.

There have been further qualitative results of interest for this model and its generalizations. The best known is that of Phelps[1973] and Sadka[1976] that the optimal marginal tax rate for the highest-wage person (if there is one) is zero. In my own paper, all wage distributions were unbounded above. The Phelps-Sadka result really says that the highest income that could possibly happen should be subject to a zero marginal tax rate. There is considerable uncertainty about the actual highest income: it is very unlikely to be close to the level at which the marginal tax would be zero.

There is an important general result due to Atkinson and Stiglitz [1976] who found nice general conditions for a model with many consumption goods to have the property that the optimum can be obtained using only a labour-income tax. This turns on separability of consumption goods in preferences from labour and consumer characteristics. If these conditions hold for intertemporal preferences, it follows that one should not have a tax on capi-

tal income: that is the case where a uniform expenditure tax is optimal. There is a closely related result of Christiansen [1981] that when there is a public good grouped with the private consumption goods, separately from labour and wage, the Samuelson public good rule, that the sum of marginal rates of substitution should equal the marginal rate of transformation, holds. These results require the possibility of arbitrary nonlinear taxation of labour income, which is perfectly reasonable. It is interesting that the general model of incentive-compatible systems gives results so much neater than those obtainable when only linear taxation is allowed.

Finally, it is worth remarking that the model is more general than it looks, for income in the model is visible income, and consumption is what the consumer is seen to be paid, net. Tax evasion can perfectly well be accommodated in the structure, with the consumption variable being apparent after-tax income, and income what is reported to the tax authority. What is missing then is other kinds of inspection and assessment. But that could be used even without deliberate tax evasion. In some countries, the possibility that evasion varies with the level of taxes is believed to be more important than variations in labour supply.

ASYMMETRIC INFORMATION

The tax model we have been considering is only one situation in which there is asymmetric information between a principal (here the government) and an agent or agents (the consumers). The individual consumer knows more about his or her own capabilities than the government. The government can think of itself as dealing with a representative consumer, but not knowing that consumer's type. Many economic relationships are of the principal/agent type, particularly employer/employee relationships. The kind of analysis I outlined applies when the agent's performance is observable and measurable, but the agent knows more than the principal, for example about the relationship between unobservable effort and performance. Adam Smith knew there was a problem (although he does not explicitly mention the uncertainties that make shirking possible):

> It is the interest of every man to live as much at his ease as he can; and if his emoluments are to be precisely the same, whether he does, or does not perform some very laborious duty, it is certainly his interest, at least as interest is vulgarly understood, either to neglect it altogether, or, if he is subject to some authority which will not suffer him to do this, to perform it in as careless and slovenly a manner as that authority will permit. (*Wealth of Nations*, V.i.f.7).

It might be thought that he is too neglectful of monetary incentive systems, but it is important to be reminded that authority could be a good description of the relationship. An optimal payment system, with asymmetric information, could well have an authoritarian character, if it showed pay rising rapidly with performance over some range, low at lower incomes, and not rising much further at higher. That would come close to the principal saying: Do

this, or else. It is an interesting feature of the schedules found to be optimal in the income-tax problem that consumption never rises more rapidly than income (equivalently, the marginal tax rate is never negative). Most probably, so harsh a relationship is never optimal in realistic cases of asymmetric information[7].

The employment relationship raises many interesting new possibilities, such as relating pay to other people's performance. That is not worth doing if the two people are quite unrelated, in that their abilities are uncorrelated. When they are correlated, and the agents cannot or do not combine together, there is indeed scope for having pay depend to some extent on relative performance.

We probably do not expect to have our government introduce taxes that depend on our neighbour's (or distant competitor's) income as well as our own. But such possibilities have played an interesting role in the further development of mechanism design beyond the simple model of asymmetric information described above. Agents can be asked to choose among much more complex sets of messages than we would use to describe their simple performance or income. Maskin[1985] introduced the idea of asking people to place themselves in the overall wage distribution, while faced with incentives that punish severely any inconsistency in the answers. By that device he was able, in a sense, to implement a first-best optimum. That theory appeared terribly demanding of information among agents, but Piketty[1993] has developed more plausible ways of getting the first best within the same general set of ideas. The simple model of incentive compatibility by no means exhausts the possible incentive mechanisms in situations of asymmetric information.

Among the other fields of application for asymmetric-information are the control of firms by regulators, and pricing by utilities. In a very interesting line of development, first Baron and Myerson[1982], and then Laffont and Tirole[1993] have shown how one can analyse regulation by treating the firm as an agent who knows its cost structure, and the regulator as a principal who is uncertain about the firm's costs. The firm's outputs and the prices it charges for them are related by market demand, and they are public information. One way of thinking about regulation is to have an output-variable tax. This can be analyzed using the methods I have described.

Similarly, utilities face consumers with varied preferences, and can relate price to quantity used in complex ways. As with the regulatory model, there are multiple products, and one should consider the simultaneous pricing of consumption at different times. This gets really hard when consumers tastes vary in a multi-dimensional way. The simple techniques I have described are much less effective in such multi-dimensional problems, but Wilson[1993] and Armstrong[1996] have made significant progress in solving problems of this type.

[7] Some of this is explored in Mirrlees [1976].

In all of these areas of application, the time dimension is potentially important. Going back to the taxation problem, it can be seen that some new and awkward problems arise. We can think of taxing each generation, or cohort, in the way appropriate to them. Each will contribute to the common pool, but we can clearly identify the year of birth, and use that as a tax base. Theoretically, therefore, we should consider having a different tax system for each cohort. Governments do not do that, and I will come back to why and whether we would like them to.

It is to be presumed that each individual's ability is quite strongly correlated with future ability. For simplicity consider a model where everyone's wage remains the same throughout the working life. Some particular incentive-compatible tax system applies to first-year income. People decide how hard to work, what to work at: some get high incomes, some low. If the theory already developed applies, people with higher wages will choose to earn higher incomes. Next year, the government knows what incomes they earned last year, and can therefore deduce their wage rates. Now it can tax on the basis of the wage rate rather than income, that is to say on the basic characteristic of the consumer. There is no longer any need to worry about incentives, not at least in the present period. Tax can be made independent of income actually earned, and related simply to the wage (observed on the basis of performance in the previous period). On the margin, incentives are optimal. In effect the wage-related taxes implement lump-sum taxes, and would be expected to be high for high wage people, low, indeed negative, for low wage people. In fact it turns out that in reasonable models, low wage people will be better off than high wage ones.

If that is going to happen in the second year, people in the first year will probably decide they would rather not earn enough to be labelled high wage. They could well all choose the same income, say zero. The second-year government cannot work out wages after all, and everything collapses. The hoped-for optimum described is not an equilibrium. There is an equilibrium with no-one doing anything, but it is extremely unsatisfactory. The trouble is that in the second year, it will be rational for the government to act as described. It is the anticipation of that rational behaviour by government that causes the trouble. What we have is a particularly bad case of intertemporal inconsistency. If the government can commit itself in advance to the tax system that will apply to the cohort in all future years, we can get back to the "second-best" equilibrium already described as an optimal tax system. Probably it can do rather better than that. The puzzle is that governments do not, to any significant extent, commit themselves to future tax rates, and indeed cannot easily do so; and yet the problem described does not arise. By accepting a convention that people born at different times are all subject to the same tax system, the government may be taken to provide such a guarantee, at the cost of giving up a desirable basis for tax discrimination.

Perhaps one should not mention it out loud. Like the man who starts thinking about how he manages to walk, we may get ourselves in a lot of trouble by thinking. We have the same problem that Kydland and Prescott [1977]

identified in macroeconomic policy. In microeconomic policy, we do not take it seriously. There is trouble lurking here, perhaps in the area of capital taxation.

MORAL HAZARD

In some degree, individual economic agents are also uncertain about their tastes and abilities when they make decisions. The extreme case, where the agent is no better informed than the principal, is well known in insurance as moral hazard. If the agent's behaviour is unobservable, it is usually not possible to deduce the individual's action from performance when the connection between action and performance is uncertain. There are many relationships where this better describes the situation than does the asymmetric information model. Medical care has been regarded as a prime example in the economics literature[8], perhaps surprisingly. Sharecropping with farmers paying for the use of land with a share of profits or income seems a good example, and so are many cases of accident insurance. Usually, of course, there are elements of each. I shall come to that at the end of the lecture.

It is interesting to examine the consumption/income model as though it were a moral-hazard model[9]. Suppose then that effective labour supply decisions are taken early in life, decisions how hard to work in school or career choices. The consequences are uncertain. In the pure moral-hazard model, everyone is identical at the point of decision. The government has to devise a tax schedule that will induce people to work or try hard at that early stage, presumably making rewards increase with income, so that these prospects will encourage early effort. It might (in simple cases it would) have wanted everyone to have the same income, but then there would be no incentive to try hard initially.

Problems of this kind are usually analysed with the assumption that people try to maximise their expected utility. There are some good reasons for thinking that may be a mistake. At least the consequences of alternative theories of decisions under uncertainty for these situations should be explored. But I shall go on with the conventional theory.

Some particular level of effort is optimal. Incentives will have to be set up so that people will do it. If this is a nicely behaved problem, and in simple cases it is, we have to arrange that the marginal effect of effort on the expected utility of consumption takes some particular value. At the same time, the government is constrained by the total consumption that is going to be available when people do that amount of effort. Subject to these two constraints, it wants to maximise expected utility. To get incentives right, some consumption levels will be low, presumably at lower income levels. At lower income levels, more effort will reduce the probability of that outcome. Reducing consumption improves incentives, though it lowers expected utility.

[8] See Arrow [1963], who has something to say about asymmetric information too; and Pauly [1968].

[9] Such a model has been examined by Varian [1980], but my discussion will be primarily a translation to the present context of part of Mirrlees [1974].

It follows that consumption should be lowered most at income levels where the incentive improvement is greatest, relative to the utility loss, i.e. where the elasticity of output probability with respect to effort is great. A relative simple idea then: reduce consumption at observed output levels where effort has a large proportional effect on the probability of that level.

There is a striking feature of the lognormal distribution of incomes for this model. If that is the nature of uncertainty about the effect of effort on income, the elasticity of probability with respect to effort tends to infinity as income goes to zero. Therefore, in the model, the government can achieve a very satisfactory outcome, almost as though the incentive constraint could be ignored, and it does that by instituting very low consumption at very low incomes. How much it can achieve that way depends on how low utility goes as consumption gets very small.

This is very peculiar, and of course unacceptable. In this particular model, one reason why it is unacceptable is that people can in fact change their labour supply "at the last minute". Another is that it assumes people can calculate intelligently about events of very small probability, which is surely not always the case. Finally, the assumption that all kinds of effort to avoid very low-income outcomes automatically increases the probability of high-income outcomes is not realistic.

Yet the analysis is trying to tell us something, something rather paradoxical. It is saying that incentives by means of punishments, which is how we might describe very low consumption when there is a very poor outcome, are most appropriate, if at all, in principal/agent situations with moral hazard. These are cases where the agent does not know the consequences of actions. More precisely, punishment may be appropriate where actions have very uncertain consequences, spread over all possible outcomes. In the opposite case, where the agent knows very well the consequences of actions, punishment is not appropriate. It just might be best to have draconian punishments for serious car accidents, but not for deliberate crime. On the whole, I do not persuade myself that solution is correct, but much remains to be done to reformulate the model so as to remove the most extreme features of the solution to what is apparently the most straightforward and natural case.

There is one feature of these solutions that is persuasive. In cases such as employment relationships where it is not possible to impose any very great punishment on the agent, dismissal will happen in a range of bad outcomes, and then rewards may rise rather rapidly over a range of outcomes. In such cases, one is getting a result not unlike a relationship of authority, where an order is given, and expected to be obeyed. Such a solution is illustrated in Figure 4.

Often one gets perfectly reasonable payment schedules in such problems, with payment, or rather the marginal utility of payment, related to the elasticity of probability with respect to effort. This gives some impression of the shape of the payment schedule, but exact computations are somewhat troublesome. The method of solution indicated does not always work, however. There are cases, and they are not at all special, where one cannot well de-

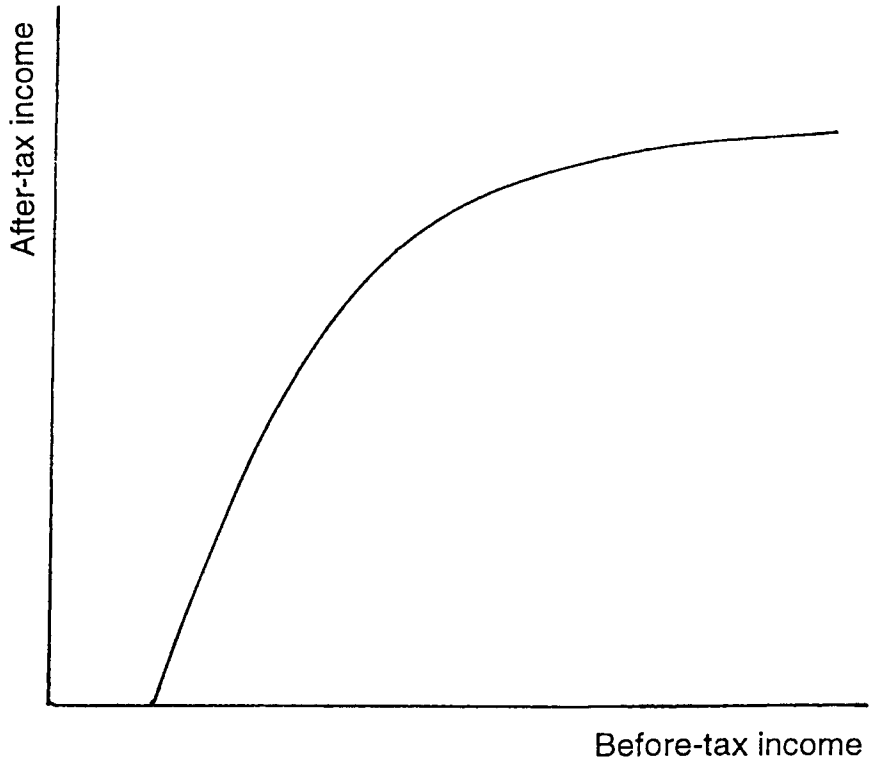


Figure 4. Optimum pay schedule under moral hazard.

scribe the moral hazard constraint, that is the constraint that incentives induce the desired level of effort, as a first-order condition the agent's choice must satisfy[10]. Sometimes the agent should be made indifferent among two or more alternatives, and induced to choose just one. For example, in a model of optimal retirement, studied by Peter Diamond and me [1977], it is a main feature of the optimum pension system that the agent, who is subject to the random onset of disability, is made indifferent as to the date of retirement, although only one particular retirement date is the right one. A very slight perturbation of the optimum schedule can induce the agent to choose that right retirement age for sure.

It is not always very easy to tell in advance what models can properly be solved by using the agent's first-order condition as a constraint. A set of sufficient conditions I had conjectured was shown to be valid by Rogerson [1985]. Surely more general conditions can be found. We have not come close to identifying the boundaries between the different cases. It is striking that in the asymmetric-information model the method of taking as the incentive-compatibility constraint essentially the first-order (calculus) conditions for choice by the agent or agents works so well, at least in the sense that a simple,

[10] Mirrlees [1975].

manageable and understandable condition is enough to justify it. When moral hazard is present, a different approach is needed.

These complications make it particularly hard to give any general rules as to what we would want or expect payment schedules to look like when the relationship between principal and agent involves moral hazard. The sharply rising middle section of the payment curve that I have referred to is by no means universal. Yet simple sharing rules are surprisingly difficult to generate with plausible examples. It is all the more striking that Holmström and Milgrom [1987] have found an example with continuous action over time where incentives are linear.

IMPERFECT INFORMATION IN GENERAL

Although one should always seek simplicity, it is necessary to consider what happens when there is asymmetric information, with the principal knowing less about the relationship between input and output than the agent who chooses input, but the agent also being in some ignorance. As the theory of these relationships was developed in the seventies, it was found that new forms of contract arose. For concreteness, let us talk about the consumption/income model again. Each consumer knows sor ething about the effect of his labour on his income. It is different for different people, and no one knows it for sure. When this is the case, there is a natural two-period structure. First the consumer chooses what to do; then, in the next period, the outcome becomes known.

In these circumstances it is almost always worthwhile for the principal to offer a choice of payment schedules to consumers, a choice that is to be exercised before income is generated, indeed before the consumer knows what income he will get as a result of his labour decisions. There is one schedule designed, perhaps one should say destined, for each type of consumer. Each consumer chooses the one that suits best. Unfortunately, the only examples that are fairly easy to work out are those where each agent's actions have a full range of possible effects, and consequently it is optimal to have a punishment schedule for each type of consumer. Maybe that is some kind of approximation to the optimal set of contracts between principal and agent, but the model is not very believable. A better model, with consumers making a succession of labour supply decisions, needs to be worked out. It is unlikely that governments will adopt such complex tax systems, or other principals impose such incentive systems on their agents, but we ought to get some sense of what the systems would look like, and by how much such possibilities might improve outcomes, both for principals and agents.

The sense that the degree of complexity implied by optimal design of incentive systems is unacceptable and unusable becomes even clearer if we make another step towards realism, and allow that consumers take a series of labour supply decisions over time: education, career choice, early efforts to achieve promotion, job search, practice and exercise, hours of work, years of work. There is some uncertainty about the consequences at each stage,

sometimes great, sometimes quite small. The formal structure of the model implies that agents should be offered not just choices among schedules, but choices among sets of schedules, among which choices will later be made. Pay or tax systems of that complexity are not conceivable. Why not? You could say that humans are not intelligent enough to take decisions of the complexity required. More reasonably, it would be very costly to calculate the decisions, and it undesirable to impose these decision costs on people. Part, perhaps a substantial part, of that cost is the cost that results from making mistakes. Mistakes are not part of the standard economic models.

Simplicity of contracts and systems is a slippery concept indeed. To recognize the desirability of simplicity is not at all to conclude that simplest is best. There may well be many equally good ways of being simple. In the area of incentive systems, it is nice linear contracts that seem to be simple. For example, it is, and has long been, tempting to conclude that a constant marginal tax rate, the same for all incomes, would do perfectly well. Some at least of the calculations that I and others have done for the asymmetric-information problem suggest that there would not be a great cost in adopting such a system, as compared with an optimal one. At least it might do just as well as any of the tax systems we have in our various countries. It has the appeal of neatness and elegance. The simplicity of so simple a tax system is not, I think, a great advantage over the slightly greater complexity of varying tax rates. The question is whether it would be much worse than an optimal tax system, and that deserves to be estimated.

But there are kinds of complexity that should still be studied. One particular example is the possibility of relating taxation to wage rates, rather than only to income. I have touched on that obliquely several times already. It is not at all as easy to study as it may seem. The earlier discussion was couched in terms of wage rates, because they are concrete and readily understandable. Formally, the definition I was using was the ratio of income to a measure of the effort an individual is making. If effort were hours of work, it would be the wage rate in the usual sense. Often effort cannot be measured that way. Using the conventionally measured wage rate, income divided by hours, as an element of the tax base would not, I suppose, add much to the use of income. It is therefore unlikely that there is much advantage in so extending the tax system. What is really wanted is taxation in relation to occupation, to the type of job, perhaps, as well as income. I do not at all know what such a optimal tax system would look like. It would be worth investigating.

Another interesting set of questions concerns taxation at the upper end of the income scale, where the relationship between pay and productivity is often far from simple. Two examples come to mind, both in need of serious analysis. The first is managers whose pay comes from contracts arising from principal/agent relationships within the firm. Because the contract of each manager will not in each year equate pay to productivity, does that mean marginal tax rates should be adjusted appropriately? The problem is made the more interesting because the form of the contract between firm and manager is influenced by the form of taxation.

The second example is that of rewards from innovation, invention, creation and competition. I might say, the question is how to treat prizes. The best singer may not be much better than the second-best, and may be primarily motivated by the wish to sing better than the second-best, or just to sing well. What does that tell us about desirable incentive contracts, and, then, about taxation at these levels? We may expect that marginal tax rates should be rather higher if high incomes are indeed generated through competitions of this kind.

Whenever one looks at a principal/agent situation, one can think of many ways in which incentives might be created. In recent years, the theory and practice of what is variously called the design of economic mechanisms, or contract theory, or principal/agent problems has gone well beyond the situations discussed in this lecture. The account given here has been one-sided, for the principal always set the terms of the contract, and the agent took the actions. It is true that many economic relationships are one-sided, in just that way. Many others are not, and involve cooperative arrangements or bargains between people in similar situations. It is not so much the asymmetry of information that is special about principal/agent relationships, but the asymmetry of responsibilities, with the principal moving first, the agent following. That makes the problems easier, and so we have made some progress. Now we can better appreciate that anonymous market relationships are only a part of economic reality; perhaps not even the largest part. Most economic problems and possibilities involve instead relationships between and among individual agents, whether taxes, contracts and bargains, fights and thefts, learning and search. It is a world still only imperfectly explored.

REFERENCES

Armstrong, C.M. [1996] Multiproduct nonlinear pricing, *Econometrica,* 64, 51–76.

Arrow, K.J. [1963] Uncertainty and the welfare economics of medical care. *American Economic Review,* 53, 941–69.

Atkinson, A.B. and Stiglitz, J.E. [1976] The design of tax structure: direct versus indirect taxation, *Journal of Public Economics,* 6, 55–75.

Baron, D. and Myerson, R. [1982] Regulating a monopolist with unknown costs. *Econometrica,* 50, 911–30.

Boiteux, M. [1956] Sur la gestion edes monoples publics astreints à l'équilibre budgétaire. *Econometrica,* 24, 22–40.

Christiansen, V. [1981] Evaluation of public projects under optimal taxation, *Review of Economic Studies,* 48, 447–457.

Diamond, P.A. and Mirrlees, J.A. [1971] Optimal taxation and public production I and II, *American Economic Review,* 61, 8–27 & 261–278.

Diamond, P.A. and Mirrlees, J.A. [1977] A model of social insurance with retirement, *Journal of Public Economics,* 10, 295–336.

Graaff, J. de V. [1957] *Theoretical Welfare Economics,* Cambridge University Press.

Holmström, B. and Milgrom, P.R. [1987] Aggregation and linearity in the provision of intertemporal incentives. *Econometrica,* 55, 303–328.

Kolm, S.-C., [1971] *La théorie des contraintes de valeur et ses applications,* Paris: C.N.R.S. Dunod.

Kydland, F.E. and Prescott E.C. [1977] Rules rather than discretion: the inconsistency of optimal plans, *Journal of Political Economy,* 85(3).

Laffont, J-J. and Tirole, J. *A Theory of Incentives in Procurement and Regulation.* Cambridge Mass.: M.I.T. Press.

Little, I.M.D. [1950] *A Critique of Welfare Economics,* Oxford: Clarendon Press.

Maskin, E. [1985] The theory of implementation in Nash equilibria, in Hurwicz, Schmeidler and Sonnenschein (eds.) *Social goals and Social Organization,* Cambridge: University Press.

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 in Equilibrium Behavior Under Uncertainty,* Amsterdam: North-Holland.

Mirrlees, J.A. [1975] The theory of moral hazard and unobservable behaviour, mimeo., Nuffield College, Oxford. To appear in *Review of Economic Studies.*

Mirrlees, J.A. [1976] the optimal structure of incentives and authority within an organization, *Bell Journal of Economics,* 7, 105–31.

Mirrlees, J.A. [1986] The theory of optimal taxation. In K. J. Arrow and M. D. Intriligator (eds.), *Handbook of Mathematical Economics, Volume III,* Amsterdam: North-Holland.

Mirrlees, J.A. [1995] Welfare economics and economies of scale. *Japanese Economic Review,* Vol. 46, 38–62.

Pauly, M. [1968] The economics of moral hazard: Comment. *American Economic Review,* 58, 531–6.

Phelps, E.S. [1973] The taxation of wage income for economic justice. *Quarterly Journal of Economics,* 87, 331–354.

Piketty, T. [1993] Implementation of first-best allocations via generalized tax schedules, *Journal of Economic Theory,* 61, 23–41.

Ramsey, F.P. [1927] A contribution to the theory of taxation. *Economic Journal,* 37, 47–61.

Rogerson, W. [1985] The first order approach to principal-agent problems. *Econometrica,* 53, 1357–1368.

Sadka, E. [1976] On income distribution, incentive effects and optimal income taxation. *Review of Economic Studies,* 43, 261–268.

Spence, M. [1973] Job market signalling. *Quarterly Journal of Economics,* 87, 355–374.

Tuomala, M. [1990] *Optimal Income Tax and Redistribution,* Oxford: Clarendon Press.

Varian, H.R. [1980] Redistributive taxation as social insurance, *Journal of Public Economics,* 14, 49–68.

Vickrey, W.S. [1945] Measuring marginal utility by reactions to risk. *Econometrica,* 13, 319–333.

Wilson, R.. [1993] *Nonlinear Pricing,* Oxford University Press.