James Mirrlees’ Contributions to the Theory of Information and Incentives*

Avinash Dixit
Princeton University, Princeton, NJ 08544-1021, USA

and Timothy Besley
London School of Economics, London WC2A 2AE, England

I. Introduction

The theory of information and incentives has revolutionized economics. We now understand how markets and other modes of private economic transaction, as well as public policy, are affected when different participants have different information, or when the actions of some participants are not observable to others. Mirrlees has made pathbreaking contributions — arguably the two most important contributions — to the development of these ideas and methods of analysis. His Review of Economic Studies article [1971c] on optimal nonlinear income taxation introduced a technique that evolved into the “revelation principle”, which fully characterizes all policies that are “incentive compatible”, or feasible under information asymmetry. This technique is now the workhorse model for numerous applications. His Bell Journal of Economics article [1976a] on hierarchies and organizations, and some of his important unpublished (yet widely circulated) papers referred to below, played a similar role for the theory of the design of incentives to induce effort. He identified the precise way in which observable outcomes convey probabilistic information about the underlying unobservable action, and thereby clarified how rewards or punishments based on outcomes can serve as incentives to action.

Mirrlees developed these ideas in the context of specific applications. However, they are truly fundamental to economic theory, and they have profoundly influenced subsequent thinking and research on the design of

* We are very grateful to Peter Diamond, Oliver Hart, Bengt Holmström, Kevin Roberts, Robert Wilson, and last and most important, Richard Zeckhauser, for their very detailed and helpful comments on the preliminary version, given under extreme time pressure.

Dates in brackets refer to publications by Mirrlees discussed in this paper (see Bibliography) and dates in parentheses to works by other authors listed at the end of this article.

Biographical note: James A. Mirrlees was born in 1936 in Minnigaff, Scotland. He received his M.S. in Mathematics in Edinburgh in 1957, and his Ph.D. from the University of Cambridge in 1963. He was Edgeworth Professor of Economics at Oxford University between 1969 and 1995, and currently holds the professorship of Political Economy at the University of Cambridge.

© The editors of the Scandinavian Journal of Economics 1997. Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK and 350 Main Street, Malden, MA 02148, USA.
incentives in numerous contexts. Robert Wilson has expressed this eloquently:¹ “For twenty years the general theory of taxation has been founded substantially on the seminal contribution by James Mirrlees. It is also the foundation for related topics that have become mature fields of study: the theory of regulation, price discrimination of various forms, nonlinear pricing and Ramsey pricing generally, contracting (labor contracts, principal–agent relationships, procurement, risk-sharing). Like the implications of Einstein’s theory of relativity for physics, the methodological implications of Mirrlees’ formulation and analysis infuse every topic in microeconomic theory where incentives and/or private information are relevant.”

Mirrlees is also known to a broader spectrum of the economics profession for his contributions to the theory of public finance, and to development economics. In the former area, in collaboration with Peter Diamond, he offered a thorough and definitive treatment of optimal commodity taxation. This both formalized and extended previous work by Frank Ramsey and Paul Samuelson. The Diamond–Mirrlees [1971a,b] papers also spurred a great deal of subsequent work, with the germs of ideas developed later by many others found within them.

Most importantly, they proved that even in the absence of personalized lump-sum transfers, it is generally desirable to main efficiency in production. Any necessary distortions should take the form of taxes or subsidies levied at the final stage of sales to consumers. In particular, for a small country that does not have any monopoly power in international trade, world prices constitute the correct social opportunity costs. This formed the basis of Mirrlees’ work with Ian Little in development economics, which has had a major practical influence on the way in which international development agencies conduct cost benefit analyses.

This essay focuses on Mirrlees’ work on information and incentives. We begin with a brief review of the issues, and the limitations of traditional economic theory in confronting them. We keep this part very non-technical and elementary, so that the general reader can obtain an adequate knowledge of the issues. We then discuss Mirrlees’ pioneering articles, explain how they evolved into the general methods of solution of problems of the design of incentives, and give a brief summary of later developments of the theory and its applications. Here we sketch a few technical details. We then briefly discuss Mirrlees’ other contributions.

**II. An Overview of Information Economics**

It has long been recognized that the participants in the numerous transactions that go on in any complex economy have diverse information

¹Private communication.

pertaining to these transactions and their effects. Managers know more about a firm's production than do shareholders; workers know more about their own abilities than do employers; taxpayers know more about their actual and potential income than does the government's revenue service.

Traditional economic theory argued that the price system automatically fully aggregates this dispersed information. The reasoning is as follows. Each producer or worker uses his information to calculate the action (supply or demand) that is most profitable or beneficial to himself, and these decisions collectively determine the market prices. Therefore market prices should reflect all the private information as it pertains to the scarcity of various goods and services.

However, this analysis applies only to markets with a large number of participants where the buyer and seller do not care about each other's identities. Many real-life transactions are not "arm's-length" or "anonymous" in this sense, and therefore they have features that orthodox theory cannot explain. Bilateral bargaining offers the most obvious example; the information a seller has about the quality of his product matters to the potential buyer, who has to decide whether to conclude a deal, and if so, on what terms. Similarly, an employer trying to secure more or better effort from a particular worker or groups of workers is not dealing in an anonymous labor market. And likewise, a bank that lends to a company faces the risk of default from that particular company, not from an anonymous capital market. Some of these relationships may be competitive ex ante: an employer may choose from among several job applicants and a worker may choose among several jobs; there are many potential borrowers and lenders in the capital market. But once a deal is struck, the relationship becomes bilateral and non-anonymous; therefore the parties must take into account this future fixity and its implications, even while they are considering the choice from among numerous potential partners.²

When private information bears on transactions that are not anonymous, each participant has the opportunity and the natural incentive to manipulate his information to his own advantage. A job applicant may overstate his ability or qualifications, and an intense worker when the boss is present may slack off in his absence. An insurance applicant may understate the risk he is seeking to insure; if an insured risk materializes, the claimant will try to hide his lack of precautionary measures.

² If one party must make some sunk investment ex ante, and the contract as struck ex ante is not fully enforceable by an external authority, then the other party can demand a renegotiation more favorable to itself; this problem of "opportunism" has been emphasized by Williamson (1985, Chs. 2, 7). The problem we discuss is different; it arises because informational asymmetries restrict the set of feasible contracts, even when feasible contracts are fully enforceable.
Of course the parties on the other side of these transactions have equally clear incentives to try to elicit the truth about quality or risk. One way would be to improve observability and make information symmetric. A bank that has lent to a company can obtain representation on the board of directors to monitor how the company is using the money; health or life insurance companies assess the riskiness of applicants by subjecting them to medical examinations; tax authorities conduct audits. But all such direct checks are costly and imperfect. Even after they have been used to the point where their marginal benefits no longer justify the marginal cost, information asymmetries persist, and they are often quite substantial. Therefore devices that rely on the informed parties' own incentives must be used.

The general idea is to structure the terms of the contract between the two asymmetrically informed parties in such a way that concealment becomes less possible or more costly for the informed party. In other words, that party's incentives must be realigned so that they are more congruent. In the context of insurance, the most obvious such devices are deductibles and coinsurance, where the insured has to bear a part of the loss; see Pauly (1968). Further analysis reveals many other, more subtle, mechanisms of this kind that are useful for more complex situations. These are discussed in Sections III and IV below. Mirrlees' work has helped lay the foundations of this general theory of the design of incentives.

The emphasis on the lack of anonymity in transactions where private information is important, and on the resulting opportunities for strategic manipulation of information, should also make it clear that concepts from non-cooperative game theory should play a central role in the theory of information economics. Indeed, it is just this combination of game theory and the theory of information that has produced some of the most exciting and fruitful research in economics during the last two decades. The prizes awarded to three pioneers of game theory — Nash, Harsanyi, and Selten — in 1994, and to two pioneers of information economics this year — Mirrlees and Vickrey — are just and fitting testimony to the importance of this research. Indeed, Harsanyi and Selten's award-winning work was to extend and refine the concept of Nash equilibrium to games of imperfect and incomplete information; this provides much of the game-theoretic basis for the modern developments of the theory of mechanism design that stem from the work of Mirrlees and Vickrey.

A Brief History

When economic theorists began to examine the role of information in detail, the first approach was to think of information as an economic commodity. Thus statistical theories of the value of information would

enable us to find the demand for information. Some useful progress along this road was made, but a different route proved much more fruitful. Instead of regarding information as a commodity by itself, people asked how the availability and the dispersal of information pertaining to ordinary economic commodities would alter the working of the markets and other allocation mechanisms for those commodities.

Arrow (1963, 1970) organized the economic analysis of asymmetric information in two categories: unobservable action (moral hazard in the jargon of the insurance literature) and hidden information (adverse selection in the same jargon). Both arise because the insured has more information about the risks he seeks to insure than does the insurance company. Moral hazard arises when the insured can take actions (such as always locking the car) that will reduce the levels of risk. The insurance company cannot observe these actions, and the availability of insurance weakens the insured’s incentives to take the precautions. Adverse selection arises when the risk depends on some innate characteristic (such as a health condition or general carelessness) that, although not under control of the insured, is known to him but not to the insurance company. Then a contract with any given ratio of premiums to coverage will attract exactly those clients whose risk is high enough that they find the ratio attractive; hence the terminology. Arrow’s analysis of these problems was mostly informal, but it is a tribute to his insight that these two concepts continue to provide an organizing framework for the subject.3

The next best known early work is Akerlof (1971). He showed how a market could collapse altogether under the weight of the adverse selection problem. The seller of a used car knows the quality of the particular car he is selling; buyers know only the average quality on the market, not the quality of a particular car offered to them. Buyers will not pay more than what the average car on the market is worth, but then only those with below-average cars will put them on the market, which will lower the average, and so on, until no one is willing to sell and the market collapses. This was a dramatic and provocative example, and spurred much further research, but by itself it is ultimately of somewhat limited value. Many markets suffer from some adverse selection, but most of them do not collapse. They find ways of coping, by devising contracts and other incentives that partially solve the problems posed by the asymmetric information. Therefore it is far more important to understand how such coping takes place, namely how to design such contracts and incentives that reveal some or all of the information, usually at some cost.

3The terminology is somewhat unfortunate; no moral judgement need be made when someone takes an unobservable action that optimizes his own objective, and selection need not always be adverse. But the usage has become common, and we will adopt it here.

There are two distinct but related ways of thinking about the resulting theory. One considers the context of the information asymmetry, distinguishing market interactions from non-market ones. The other, conceptually more basic, classifies the problems according to whether the initiative for coping comes from the less informed party or the more informed.

Methods that the less informed party can employ to elicit desired information or action from the more informed party are generally called “screening” devices. Direct tests are screening by examination; schemes that alter the informed party’s own incentives in the desired direction are called screening by self-selection. Mirrlees’ and Vickrey’s work provides foundations of the latter branch of theory. It is of particular importance in non-market situations — the internal organization of a firm, and the government’s economic policy and planning. Of course screening by self-selection does arise in market contexts too; an important early model for insurance markets is in Rothschild and Stiglitz (1976) and Wilson (1977).

The more informed party may wish to take the initiative and reveal its information if this will work to its advantage, but then the other party will treat such revelations skeptically. A highly skilled or productive worker would like a potential employer to know the fact; so would a good car driver like the insurance company to recognize his skill. The difficulty is that a less skilled or productive worker would try to pretend to have high skills or productivity, and a worse driver would claim to be better. Therefore strategically aware employers or insurers would not believe mere assertions. The truly skilled or careful have to find observable actions that those with less skill or care would find it too costly or impossible to imitate, and which therefore credibly convey their truthful information to the less informed party in the transaction. Spence (1974), who pioneered this branch of the theory of asymmetric information, labelled such actions “signals”. For example, a high-skill worker may get education to a level that someone with less skill would find it too expensive to achieve. Even if the last years of education have no direct effect on the worker’s productivity, they can still have value as a signal. Signalling is particularly important in market interactions, although it does arise in organization or planning, too.

This very brief discussion of signalling merely serves to clarify how that branch of the theory differs from the one where the uninformed party takes the initiative, namely screening, which is our main focus here. We now proceed to examine in detail the design of incentive mechanisms for screening, first for the case where some innate characteristic of one party is not known to the other (adverse selection), and then for the case where one party’s action is not observable to the other (moral hazard).
III. Incentives under Adverse Selection

To recapitulate, the concept of adverse selection in information economics arises most directly in an insurance context. If some innate characteristic that affects the risk class of an individual is not observable to the insurance company, then an insurance policy will selectively attract clients in bad risk classes. More generally, the term refers to a situation where one party to an economic relationship has advance private information about some parameter that affects the payoffs from the relationship.

An important application of adverse selection is the case of redistributive taxation where the government cannot observe individuals' earnings abilities. Mirrlees [1971c] income tax model was the first general analysis of optimal policy design that took account of such informational asymmetry. Edgeworth (1897) had observed that an egalitarian objective implies very high tax rates — even 100 per cent at the margin. Although economists were aware of the implied disincentives, little progress had been made on how to bring these into the analysis, weighing up the gains from equity against any efficiency loss. Vickrey (1945) made a valiant attempt to model this, but did not get any results. There were attempts that specified a parametrized tax schedule, for example affine in Sheshinski (1972), quadratic in Zeckhauser (1969), or isoelastic in Wesson (1972), but these remained isolated examples that failed to bring out the general principles. Mirrlees provided a formulation.

The following is a somewhat simplified version of his model. Let $x$ denote the consumption or after-tax income of a person, and $z$ the output or before-tax income he produces. People differ in their productive efficiency. These differences are captured by an index $n$ over a continuum range $[l, h]$. Normalize the total population to unity. Let $f(n)$ denote the density function of the population according to their types, and $F(n)$ the cumulative distribution function of types. A person with higher $n$ can produce a given output $z$ with less effort or disutility. Choose the scale of $n$ so that the utility of person $n$ is $U(x, z/n)$, increasing in the first argument, decreasing in the second, and concave. A simple interpretation is that by working $y$ clock hours, person $n$ can generate $z = ny$ efficiency units of labor.

The government wants to maximize a standard Utilitarian welfare function

$$
\int_{l}^{h} U \left( x(n), \frac{z(n)}{n} \right) f(n) \, dn. \tag{1}
$$

This is subject to the budget constraint, or equivalently a total resource constraint,
Integrating, \[ \int_t^h [z(n) - x(n)] f(n) \, dn \geq 0. \] (2)

If each person's type were publicly observable, the government could impose individualized lump-sum taxes. This is formally equivalent to ordering person \( n \) to produce output \( z(n) \), and give him consumption \( x(n) \), these functions being chosen to maximize (1) subject to (2). The first-order conditions are
\[ U_1\left(x(n), \frac{z(n)}{n}\right) = \lambda = -\frac{1}{n} \cdot U_2\left(x(n), \frac{z(n)}{n}\right), \] (3)

where \( \lambda \) is the Lagrange multiplier for the constraint (2). This is a hypothetical first best, assuming away information problems. It is essentially the solution that Edgeworth had in mind, since the marginal utility of consumption is equalized across individuals.

To see why such information problems are endemic, differentiate (3) totally with respect to \( n \), solve for the derivatives \( x'(n) \) and \( y'(n) \), and substitute to find the effect on the utility \( V(n) = U(x(n), z(n)/n) \). As long as leisure is a normal good, this implies \( V'(n) < 0 \), so that high ability individuals are worse off than those of low ability. Since the government wants to equalize marginal utilities of consumption and effort, the most productive individuals must work harder, while not enjoying sufficiently increased consumption to compensate them.

This starkly highlights the incentive problem: if \( n \) is not publicly observable, and the government tries to implement the above first-best, people will have an incentive to pretend to be less productive than they actually are by reducing their labor supply. This makes the first best infeasible.

The central question then becomes what kind of a second best can be achieved using an income tax schedule common to all persons. Such a schedule can be equivalently specified by making the consumption or after-tax income \( x \) a function of the production or pre-tax income \( z \), say \( x = X(z) \). Each person \( n \) chooses \( z(n) \) to maximize utility \( U(x, z/n) \) subject to \( x = X(z) \).

The schedule is chosen by the government. Properties of the tax schedule, such as progresivity, must be derived from the government's optimization; they cannot be assumed from the outset. Therefore virtually none of the usual convexity conditions for the consumer's choice can be imposed. This makes the problem seem intractable. Mirrlees' ingenious device was that under certain conditions the whole problem could be reduced to just one optimization — the government's — with a subsidiary constraint to ensure that the \( z(n) \) in the government's solution would also be chosen by person \( n \).
We shall state Mirrlees’ theorem, and then in the subsections that follow, relate it to later developments in this area. Suppose that the function
\[ \phi(x, y) = -yU_2(x, y)/U_1(x, y) \] (4)
is increasing in \( y \) for each fixed \( x \). The consumption, production and utility levels \( x(n) \), \( z(n) \) and \( V(n) = U(x(n), z(n)/n) \) over the range of \( n \) arise from individual maximization under some tax function \( x = \chi(z) \) if and only if two conditions hold: (i) \( z(n) \) is a non-decreasing function, and (ii) utility varies with individual type according to the differential equation
\[ V'(n) = -z(n)U_2(x(n), z(n)/n)/n^2. \] (5)

This theorem reduces the very difficult problem of designing the schedule \( \chi(z) \) to a simpler and more standard optimal control problem, namely the maximization of (1) subject to the constraint (5). Regard \( n \) as the independent variable, \( V \) as the state variable, and \( z \) as the control variable; given \( V \) and \( z, x \) is solved using the relation \( V = U(x, z/n) \). Then (5) becomes the equation for the system dynamics. Note that since \( U_2(') \) is negative, \( V'(n) \) is positive. Thus the higher types must be given sufficiently more utility to remove their incentive to pretend to be lower types. This control problem can be solved whereas the original one degenerates into a morass of uninterpretable conditions. This transformation was Mirrlees’ breakthrough.5

Mirrlees’ equivalence theorem contains within it the seeds of much subsequent work on mechanism design with hidden information. The next two subsections explain this by relating the theorem to two lines of generalization that have followed.

The Revelation Principle

The most important innovation in Mirrlees’ solution is that instead of the income tax schedule the government can specify the consumption and output schedules \( x(n) \) and \( z(n) \) subject to certain conditions. The differential equation (5) captures the condition that the assigned quantities should be equivalent to the ones individuals would choose optimally along a tax

4 Actually, our statement of the theorem ignores several subtle matters to do with the definition of the consumption set, the differentiability of \( V(n) \), etc. But the simplified version suffices to bring out all the points that are relevant here.

5 More generally, this is known as solving the problem by focusing only on “local” incentive compatibility conditions which concern only whether a particular individual wishes to masquerade as an adjacent type. In this analysis, local incentive compatibility implies global incentive compatibility, i.e., not wishing to masquerade as an adjacent type guarantees that one would not wish to masquerade as any other type. See the discussion below.
schedule. Since person \( n \) could have made the choice \((x(m), z(m))\), we must have

\[
U(x(m), z(m)/n) - U(x(n), z(n)/n) \leq 0. \tag{6}
\]

Omitting some technical problems of discontinuities for ease of exposition, suppose that \((x(n), z(n))\) varies smoothly with \( n \) along the schedule. Divide (6) by \((m-n)\) and let \( m \to n \). This gives

\[
\frac{U_1(x(n), z(n)/n)}{n} x'(n) + \frac{1}{n} U_2(x(n), z(n)/n) z'(n) = 0. \tag{7}
\]

Then

\[
V'(n) = \frac{dU(x(n), z(n)/n)}{dn}
\]

\[
= U_1(x(n), z(n)/n) x'(n) + \frac{1}{n} U_2(x(n), z(n)/n) z'(n) - \frac{z(n)}{n^2} U_2(x(n), z(n)/n).
\]

Using (7) in this, we get (5); this is just an “envelope” result.

All this works as if the government simply asked people to declare their private information (their type \( n \)) and then assigned them appropriate production targets and consumption levels. The functions \( x(n), z(n) \) are laid down in advance; they constitute the government’s commitment about how it will use the information that people supply. Then (6) becomes the condition that the functions should be such that people find it optimal to reveal their information truthfully. Once the resulting control problem is solved, the optimal policy can be expressed as a tax schedule \( x = X(z) \) obtained by eliminating \( n \) between the optimal control functions \( x = x(n) \) and \( z = z(n) \).

These ideas were later formalized by others, most notably Myerson (1982), into some terminology and a theorem. A mechanism that simply asks people to reveal their information is called “direct”; if truthful revelation is optimal, such a mechanism is called “incentive compatible”. Then the general theorem, labelled the revelation principle, says that all feasible allocations of any mechanism are the same as those of a direct incentive-compatible mechanism.\(^6\) This principle dramatically simplified the problem by permitting us to restrict our search for an optimal

---

\(^6\)Once again we leave out many details, most importantly the specification of equilibrium concept under each mechanism. In the Mirrlees model, we can use the most compelling concept, namely dominant strategy equilibria.

mechanism without loss of generality to a much smaller class of feasible mechanisms.

*The Single-Crossing Property*

The simplification afforded by the revelation principle goes only part of the way. The complete or global incentive compatibility condition is (6). It has to hold for all \( n \) and \( m \). This is still a continuum of constraints; maximizing (1) subject to all of them at once would be too complex. The further trick in reducing the problem to a standard control framework is the ability to replace this set of conditions by the local or first-order conditions (7), or equivalently, the envelope result (5). That is achieved by the condition on the function \( \phi(x, y) \) defined in (4) above.

For a fixed \( n \), consider the tradeoff between \( x \) and \( z \) along an indifference curve. The implicit function theorem gives

\[
\frac{dx}{dz}_{U = \text{constant}} = -\frac{U_z(x, z/n)(1/n)}{U_1(x, z/n)} = \frac{1}{z} \phi(x, y),
\]

where \( y = z/n \). Note that the curves are upward sloping because \( z \) is a "bad"; they are convex because \( U \) is quasi-concave. Now compare the slopes of the indifference curves for different \( n \) that intersect at a given point \((z, x)\). A higher \( n \) with given \( z \) means a lower \( y \). Since \( \phi(x, y) \) has been assumed to increase as \( y \) increases for a given \( x \), (8) shows that of the two individuals whose indifference curves cross through the point \((z, x)\), the one with the higher \( n \) has the flatter slope. In particular, any pair of indifference curves for different \( n \) can cross only once; hence the name *single-crossing property*.

This assumption ensures that the functions \( x(n) \) and \( z(n) \) obtained using the first-order envelope property in (5) do achieve the desired individual maxima for all \( n \). In terms of a tax schedule, the idea is that higher-\( n \) types would find their optimum farther to the right, and thus would choose a larger \( z(n) \). We omit the details.

The single-crossing property underlies virtually all models of adverse selection, and Mirrlees was the first to recognize it and pinpoint its role. The work of Spence (1974), Rothschild and Stiglitz (1976) and others further clarified its role, and popularized its use.

Single-crossing allows us to replace the global incentive-compatibility condition (6) by the local (7). But it also implies that higher-\( n \) individuals will choose a larger \( z(n) \); therefore the government’s optimization must
recognize this as a constraint on feasible or implementable functions \((x(n), z(n))\). This needs some care. The usual starting point is to ignore this condition and see if the solution automatically generates an increasing \(z(n)\). If it does not, then the range of \(n\) has to be split into intervals where \(z(n)\) is increasing and ones where it is constant. This is called the “ironing procedure”; see Baron and Myerson (1982) and Wilson (1993, p. 85) for details.

The Optimality Conditions

Even with the important conceptual simplification afforded by the revelation principle, the optimal income tax model remains quite complicated, and Mirrlees’ [1971c] solution of it required much additional skill and effort, of both economic and mathematical varieties. Here we want to elucidate the issue of incentives in the simplest possible way, so we examine a very special case. We assume a utility function linear in consumption:

\[
U(x, y) = x - C(y)
\]

where the disutility \(C(y)\) of the clock hours \(y\) is an increasing and convex function. This is a poor model for the income tax problem as such, because in conjunction with the utilitarian objective (1) it eliminates the equity considerations that motivate the problem in the first place. The important compensating advantage for us is that the aspect of incentives stands out most clearly and simply, and brings out the close parallel with other frequent uses of incentive compatible mechanisms as in industrial organization where income effects can be neglected; see Fudenberg and Tirole (1991, pp. 262–5).

To restore any role for any income tax, we suppose that the government has a fixed revenue requirement \(R\); thus the tax schedule is to be designed to raise this revenue while imposing as small a dead-weight burden as possible. There is one further problem. When utility is linear in consumption and social welfare does not value equality, incentive compatibility can be met costlessly by driving down the utility of the low types very far and then giving sufficiently more to each higher type. Therefore we suppose that there is some lower bound on consumption which becomes binding for a range of the lowest types. We do not treat this in detail below, but focus on the more interesting first-order conditions for the types who are above this floor level.

---

7 Subsequent work has shown that this holds in many applications if the “hazard rate” of the distribution function of types, \(f(n)/(1- F(n))\), is a monotonic function of \(n\).
Regarding the utility $V(n)$ as the state variable, and formally taking $y(n)$ to be a control variable instead of $z(n)$, the other control variable $x(n)$ is linked to these by

$$V(n) = x(n) - C(y(n)). \quad (9)$$

Then the budget constraint (2) becomes

$$\int_l^h [ny(n) - C(y(n)) - y(n)] f(n) \, dn = R. \quad (10)$$

The incentive compatibility condition is

$$V'(n) = \frac{y(n)C'(y(n))}{n}. \quad (11)$$

Subject to these conditions, we wish to maximize,

$$\int_l^h V(n)f(n) \, dn. \quad (12)$$

While this can be solved using Pontryagin’s maximum principle, it is even easier to think of it as an ordinary Lagrangian problem with a continuum of choice variables $x(n)$ and $V(n)$, subject to a single revenue constraint (10) and a continuum of local incentive compatibility conditions (11). Let $\lambda$ denote the Lagrange multiplier for the former and $\psi(n)$ those for the latter. Then the Lagrangian is

$$L = \int_l^h \{V(n) - \lambda[ny(n) - C(y(n)) - V(n)]\} f(n) \, dn + \psi(h)V(h) - \psi(l)V(l)$$

$$- \int_l^h \psi'(n)V(n) \, dn - \int_l^h \psi(n)[y(n)/n]C'(y(n)) \, dn,$$

where in the second line we have integrated the product $\psi(n)V'(n)$ by parts.

The first-order conditions for $V(n)$ are

$$(1 - \lambda)f(n) - \psi'(n) = 0.$$

Since $V(h)$ is unconstrained, the terminal or transversality condition at the upper end-point $h$ is $\psi(h) = 0$. Then we have

$$\psi(n) = (\lambda - 1)[1 - F(n)]. \quad (13)$$

If the revenue constraint has bite, relaxing it should make it possible to increase social welfare, which is being measured in consumption units, more than one-for-one. Therefore we may take $\lambda > 1$. 

Next, the first-order conditions for $y(n)$ are
\[
\lambda[n - C'(y(n))] f(n) - \psi(n) [C'(y(n)) + y(n)C''(y(n))]/n = 0. \tag{14}
\]
These have a very intuitive and useful interpretation, which we show using Figure 1. We violate the mathematical formalism somewhat, and suppose that the people are located at close but discrete points $n, n + dn, \ldots$, the number at $n$ is $f(n)\,dn$, and the numbers at $n + dn$ and beyond total $[1 - F(n)]$. In the figure, type $n$ is located at point $N$, and type $(n + dn)$ at point $D$. Their indifference curves are labelled $I(n)$ and $I(n + dn)$, and are shown as straight lines because this is a “local” picture covering only a small range of output and consumption quantities. The single-crossing property ensures that $I(n + dn)$ is flatter than $I(n)$. Also, $I(n + dn)$ passes through the point $N$; therefore type $(n + dn)$ is only just prevented from wanting to masquerade as $n$; this is how the incentive compatibility condition binds in the optimum solution.

Now suppose the government asks all persons of type $n$ to work $dy$ more clock hours. This raises their output by $n\,dy$. They are given enough extra

![Fig. 1.](image-url)

consumption to keep them indifferent; this moves them from point $N$ to point $N'$ in the figure. Using (8) above, the slope of the indifference curve $I(n)$ is $C'(y)/n$, so the extra consumption to each is $n \frac{dy C'(y)}{n} = C'(y) dy$. Therefore the government raises extra revenue $[n-C'(y)] dy$ from each (corresponding to the height $AN'$ in the figure). There are $f(n) \, dn$ such people, and each unit of revenue is valued at $\lambda$, so the increase in the government's objective is

$$\lambda [n-C'(y(n))] f(n) \, dn \, dy.$$ 

This is just $dn \, dy$ times the first term in (14) above.

But the action has additional consequences beyond $n$. Consider the type-$(n + dn)$ people at $D$. They can now pretend to be of type $n$, and get the higher utility at $N'$, because they can produce the required higher output $n(y(n) + dy)$ at a smaller disutility of $C(n(y(n) + dy)/(n + dn))$. To prevent this, their consumption must be raised. In the figure, they must be moved up from $D$ to $D'$. We have assumed utility to be linear in consumption, so $DD' = BN'$. The latter height equals the horizontal distance $n \, dy$ times the difference between the slopes of $I(n)$ and $I(n + dn)$, type type-$n$ and type-$(n + dn)$ indifference curves at $N$. The slope of $I(n)$ is $C'(z/n)/n$, and holding $z$ fixed at its value at the point $N$, the change in the slope as $n$ increases by $dn$ is

$$-\frac{d}{dn} \left( \frac{C'(z/n)}{n} \right) dn = -\frac{n C''(z/n) (-z/n^2) - C'(z/n)}{n^2} \, dn$$

$$= \frac{C'(y) + yC''(y)}{n^2} \, dn.$$ 

Multiplying this by $n \, dy$ gives the extra consumption required for types $(n + dn)$. That increases the utility of each person receiving it by an equal amount, but it is also a revenue loss to the government, and that is valued at $\lambda$ per unit.

Moreover, the same consumption increase must be granted (equal because utility is linear in consumption) to the people of type $(n + 2 \, dn)$ to prevent them from pretending to be type $(n + dn)$, and so on. There are in all $[1-F(n)]$ people to the right of $n$. Giving them all this extra consumption lowers the government's objective by

---

8 Remember that the clock hours are not directly observable.

9 For more general utility functions, these compensation amounts for successively higher types of people will be all different, and more complicated calculations will have to be made to determine the correct amounts and then to integrate them over types from $n$ to $h$. The resulting formula will be more complex, but the principle of solution is the same as in our special case.
\[(\lambda - 1)[1 - F(n)] \frac{C'(y(n)) + yC''(y(n))}{n} \ dy \ dn\]

which is just \(dy \ dn\) times the second term in the optimality condition (14) above.

Putting together the direct effect to do with the people at \(n\), and the indirect effect via the people to the right of \(n\), we see that the condition requires the full marginal effect to be zero. This is just as it should be.

**Further Results**

The transversality condition \(\psi(h) = 0\) has an important by-product. The first-order condition (14) at that point becomes \(h - C'(y(h)) = 0\). The most productive person is asked to work to the point where the marginal disutility equals the productivity. In other words, there is no distortion of effort at that end-point.

In the language of the tax schedule, this has a dramatic implication. Differentiating the utility relation (9) and using the incentive compatibility condition (11), we have

\[x'(n) = y'(n) + C'(y(n))y'(n)\]

\[= \frac{y(n)}{n}C'(y(n)) + C'(y(n))y'(n)\]

\[= C'(y(n))\frac{y(n) + ny'(n)}{n} = C'(y(n))z'(n)/n.\]

Along the income tax schedule \(x = X(z)\),

\[dx/dz = x'(n)/z'(n) = C'(y(n))/n.\]

Therefore at the upper end-point \(n = h\), we have \(dx/dz = 1\), so the optimal marginal tax rate at the top is zero.

This may seem strange, but it has a very intuitive interpretation that follows from our discussion of the condition (14) above. Inducing person \(n\) to work a little harder by offering more consumption has the additional bad consequence that all the people above \(n\) must also be offered higher consumption to preserve their incentive compatibility. This is irrelevant for the person at the top end of the range. Therefore it is optimal to induce him to work harder so long as any positive surplus can be extracted; this stops being desirable only when the point is reached where the marginal product of the top person's effort equals its marginal disutility.

This intuition shows that the "no-distortion-at-the-top" result is far more general than the specific context illustrated here. We ignored equity considerations by making utility linear in consumption, but the result does not depend on this simplification. If a tax schedule does not have this property, another feasible and Pareto superior schedule can be
constructed, proving the first one non-optimal; see Phelps (1973), Sadka (1976) and Seade (1977). The result has also proved very important in other applications, most particularly for non-linear pricing.

This result on the zero marginal tax rate at the top is, however, often misconstrued. First, its applicability is found to be limited in the numerical exercises discussed below. It also applies only when the support of the distribution is known for sure. Second, it is a distraction from the main issue for effecting redistribution, which concerns the average tax rate. The average tax rate for the richest person may be quite high, and rising through most of the range of \( n \) even when the marginal rate at the top is zero.

The precise assumptions about what is observable make an important difference for the incentive compatibility condition and therefore for the optimum tax schedule. Throughout the above analysis, we have assumed that the government could observe a person's gross output or income \( z \), but not the effort or clock hours \( y \). A different solution results if \( y \) can be observed. Then the government can announce functions \( x(n) \) and \( y(n) \). The truthful revelation constraints become

\[
U(x(n), y(n)) \geq U(x(m), y(m)) \quad \text{for all } n, m.
\]

These imply equal utility for all persons. The mechanism simply strings out everyone along a common indifference curve, with the higher-\( n \) types working longer hours and getting just enough more consumption to compensate them. Mirrlees [1974b] pointed this out, but also remarked that in reality clock hours were not a true indication of effort. The translation of clock hours \( y \) into effective output \( z \) required some unobservable effort, and a mechanism based on \( y \) gave no incentive to the higher-\( n \) types to make this effort.

**Subsequent Developments and Applications**

Mirrlees' pioneering work on incentive compatibility and the design of optimal reward schedules under information asymmetry has led to numerous applications, and has transformed our understanding of many issues in many fields of economics. We offer a very brief and selective summary of these subsequent developments.

**Taxation:** Mirrlees was motivated by the desire to understand properties of just tax systems. For practical application, it is important to discern properties of the tax schedule more fully, since the first-order conditions tell us very little on their own. The first set of simulation results were presented in Mirrlees [1971c]. He found the tax schedule to be roughly linear, with fairly low marginal tax rates. Stern (1976), beginning from Mirrlees' original
finding of near linearity, conducted a more comprehensive analysis of a linear tax scheme. He tried different specifications of the production function or the objective function to see how high or progressive the tax schedule could become, with mixed success. Mirrlees' simulations imposed strong assumptions about preferences and they were later augmented by the more comprehensive exercise of Tuomala (1984). He confirmed that zero at the top is a not particularly important in applications with rather high marginal rates being found near to the top of the distribution, then going rapidly to zero. He also found considerable non-linearity in the optimal tax schedule.

Hammond (1979) brought out some of the links between optimal taxation and mechanism design more clearly. He examined informational feasibility in greater depth, and established the important result that in an economy with a continuum of types, personalized lump-sum transfers are not incentive compatible. This gives a theoretical basis for the usual assertion that such transfers, despite their theoretically desirable property of being non-distorting, are not a practical tool of policy. In an economy with a finite number of discrete types, limited lump-sum transfers are feasible; Stiglitz (1981) examines this possibility and its consequences.

Atkinson and Stiglitz (1980, pp. 412--22) give a simple exposition of the theory of optimum non-linear income taxation with finite numbers as well as a continuum of types, and offer an interpretation of the condition (14) that is specifically tailored to this context. Guesnerie (1995) is a recent detailed and masterly analysis of all aspects of the theory of tax design under asymmetric information.

A further important line of research concerns when an optimal income tax constitutes the limits on the amount of redistribution that a society can undertake. Part of this issue concerns whether income taxes should be supplemented with commodity taxes and/or public provision of private goods to gain even more redistributive possibilities. Beginning with Atkinson and Stiglitz (1976) and Mirrlees [1976a], this has been shown to depend upon the separability between leisure and other goods in the utility function. Intuitively, a good that is relatively more complementary with labor supply ought to be subsidized. However, if all goods are equally substitutable or complementary with labor supply (as in the case of separability), then income taxes can serve all redistributive ends.

The Revelation Principle: While Mirrlees [1971c] already contains the idea, the generality and the power of this principle became clear only gradually in the course of research over several years. The same idea also emerged from another source, namely the analysis of manipulability of social choice rules by Gibbard (1973), which in turn grew out of a conjecture of Vickrey (1960). More general formulations, and different equilibrium concepts
(dominant strategy, Bayesian Nash) emerged with the work of Dasgupta, Hammond and Maskin (1979) and Myerson (1982).

Non-linear Pricing: Spence (1977) was probably the first to recognize the generality and wider applicability of Mirrlees’ income tax model. He adapted the technique — regarding utility as a state variable, whose differential equation captures the local incentive compatibility condition, and imposing the single crossing condition to ensure global incentive compatibility — to characterize a monopolist’s optimum non-linear price schedule when he does not know the demand function of any individual consumer. Indeed, the single-crossing property is now often called the Mirrlees–Spence condition. Mussa and Rosen (1978) in an early and related model treat the question of a multi-product monopolist’s choice of qualities and prices.

This line of research has progressed far in industrial economics. The masterly book by Wilson (1993) is a notable recent contribution.

Regulation: Baron and Myerson (1982) showed how a regulator would set prices when the costs of the firm were private information. Just as the marginal tax rate in Mirrlees’ problem has to be small enough to provide highly productive individuals to reveal their ability through their income-leisure choice, so the regulator has to allow low-cost firms sufficiently higher profit (share the surplus or rent with them) to prevent them from slackening or pretending to have higher costs. Apart from its specific contribution to the theory of regulation, this paper was important in explaining this idea to a much wider audience, and thereby spurring much more research using the revelation principle. This line of research culminated in the comprehensive treatise by Laffont and Tirole (1993), which also initiated an extension to the politics of regulation.

Public Goods and Auctions: An individual beneficiary of a public good has better knowledge of his own preferences than does the government. The consumer has the incentive to understate his willingness to pay, and the government wants to elicit the truth. The same problem arises between each bidder and the auctioneer when a private good is being auctioned. The theory of design of truth-revealing mechanisms in these situations got its start from the work of Vickrey. Later developments of this research used the revelation principle, and its associated techniques, more formally. Myerson (1981) developed and applied the principle in the context of

---

10Spence does not cite Mirrlees [1971c] directly, but he cites a manuscript version of Atkinson and Stiglitz (1980), whose discussion of income taxation is explicitly based on Mirrlees’ paper.
auctions. Green and Laffont (1979) gave a thorough treatment of the theory of mechanism design for demand revelation and efficient provision of public goods.

IV. Incentives under Moral Hazard

The essence of moral hazard is that one party to an economic relationship takes an action that cannot be observed or inferred by another affected party. Thus an observable outcome $x$ is a function $x = g(a, \theta)$ of two unobservables, an action $a$ and a state of nature $\theta$. The action is chosen before the state of nature is realized. In the context of insurance, where the term moral hazard originates, $a$ could be the care exercised by the insured, and $x$ the amount of loss in an accident. In a principal–agent relationship, $a$ is the agent’s effort and $x$ the total profit or surplus for the pair. In each case, $\theta$ is a random variable that affects the outcome, making it impossible to infer $a$ perfectly by observing $x$.

Following the general discussion of insurance by Arrow (1963), the problem was modelled in this way by Spence and Zeckhauser (1971) and Ross (1973), although the latter contribution focused on the case where effort was not costly. We lay out this formulation briefly, and then turn to Mirrlees’ reformulation.

The contract between the principal and the agent specifies the amount $s(x)$ the principal is to pay the agent if the outcome $x$ is observed. Let $H(x - s(x))$ be the principal’s utility, and $U(s(x)) - V(a)$ the agent’s, net of his disutility of effort. Let $\phi(\theta)$ denote the density function of the random variable $\theta$. The principal is to choose the contract $s(x)$ to maximize his expected utility

$$\int H(g(a, \theta) - s(g(a, \theta)))\phi(\theta) \, d\theta. \quad (15)$$

subject to two constraints. First, he must ensure that the agent has sufficient expected utility $u_0$ to induce him to accept the contract (the participation constraint)

$$\int [U(s(g(a, \theta))) - V(a)]\phi(\theta) \, d\theta \geq u_0. \quad (16)$$

Second, the principal must recognize that, given the contract $s(x)$, the agent chooses $a$ to maximize his own expected utility (the incentive compatibility constraint). Thus

$$a \in \arg \max \int [U(s(g(a, \theta))) - V(a)]\phi(\theta) \, d\theta. \quad (17)$$

In practice, this condition must be replaced by the first-order condition of the agent’s optimization. But that condition requires the derivative of the payment function \( s(x) \); by using it we are imposing differentiability on the principal’s choice function whereas any such properties, if they are valid, should emerge as a part of the solution. A second drawback of the above method is of greater practical importance: the conditions for the principal’s optimization problem become unwieldy and yield little general insight, although Spence and Zeckhauser (1971) and Ross (1973) examined several special cases and obtained some useful inferences.

This set the stage for another important breakthrough by Mirrlees. We offer a very brief exposition. The details can be found in Mirrlees [1974b, 1975a, 1976a, 1979a]; see also the survey by Hart and Holmstrom (1987).

Mirrlees’ crucial analytical advance came from the fact that for each level \( a \) of effort or care, the probability distribution of the random state \( \theta \) gives rise to a probability distribution on the outcome \( x \). Therefore the choice of \( a \) can be regarded as a choice among the available probability distributions of \( x \). Write \( F(x, a) \) for the cumulative distribution of \( x \) given \( a \), and \( f(x, a) \) for the corresponding density function. More effort produces better outcomes in the sense that as \( a \) increases, the distribution of \( x \) shifts to the right in the sense of first-order stochastic dominance.

Now the expression (15) for the principal’s expected utility gets replaced by

\[
\int H(x-s(x))f(x, a) \, dx, \tag{18}
\]

the participation constraint (16) is replaced by

\[
\int [U(s(x))-V(a)]f(x, a) \, dx \geq u_0, \tag{19}
\]

and the incentive compatibility constraint (17) becomes

\[
a \in \arg \max \int [U(s(x))-V(a)]f(x, a) \, dx. \tag{20}
\]

Note that \( a \) no longer appears as an argument of \( s(\cdot) \); therefore the condition of the optimal choice of \( a \) can be characterized without imposing

\footnote{Spence and Zeckhauser (1971, footnote 2) mention this possibility, but do not conduct any analysis using it.}
a priori assumptions on the reward function. We turn to the results and interpretations that emerge from this formulation.

The Tradeoff Between Incentives and Risk-Sharing

Suppose the unorthodox constraint (20) can be replaced by a customary equality constraint, namely the first-order condition of the agent’s maximization,

\[-V'(a) + \int [u(s(x)) - V(a)] f_a(x, a) \, dx = 0. \quad (21)\]

Now the principal wants to choose \( s(x) \) to maximize (18) subject to (20) and (21). This is a simpler variational problem, and its first-order condition is

\[
\frac{H'(x - s(x))}{U'(s(x))} = \lambda + \frac{f_a(x, a)}{f(x, a)}
\]

where \( \lambda \) and \( \mu \) are the respective Lagrange multipliers for the constraints (19) and (21). One can prove that \( \mu > 0 \); see Holmström (1979) and Roger-son (1985) for further discussion.

This has an immediate interpretation. If the incentive compatibility constraint were absent, the second term on the r.h.s. would be absent, and the ratio of the marginal utilities of income to the two parties would be equal in all outcomes. This is the Arrow–Borch risk sharing result, which would yield the standard first-best Pareto optimum of an Arrow–Debreu model of complete markets in the absence of informational asymmetry. Therefore the second term on the r.h.s. of (22) captures the extent to which risk-sharing must be limited to give the agent an incentive to exert the unobservable effort.

The existence of a tradeoff between risk and incentives was noted by earlier writers. In fact an analogous condition appears in Ross (1973, eq. (10)). But Mirrlees’ result makes the idea much more precise and amen-able to a very intuitive interpretation which is due to Holmström (1979).

Note that

\[
f_a(x, a)/f(x, a) = \partial [\ln f(x, a)]/\partial a.
\]

This is the derivative of the log likelihood function of outcome, with respect to the effort level viewed as an unknown parameter. It gives the principal probabilistic information based on the observable \( x \) about
whether the agent took the unobservable action $a$. If the derivative is large and positive, the principal can quite reliably infer that the agent’s action was unlikely to be locally smaller than $a$. Therefore he can more confidently reward the agent based on this $x$. That is just what (22) shows. The larger is the right-hand side, the larger is the principal’s marginal utility relative to that of the agent, and therefore the larger the agent’s reward $s(x)$ from the outcome $x$. In other words, the more informative is $x$ about $a$, the farther will the incentive payment for $x$ depart from that under first-best risk-sharing.

This elegantly simple result not only gave economists a clear intuition about the trade-off between risk-sharing and incentives, but also paved the way for further understanding of the role of other observable and informative variables in contract design. At this level of generality, one cannot say much about the shape of the optimal incentive scheme. However, monotonicity in $x$ such that high observed returns are associated with a higher payoff would appear to be a natural property. This will hold provided that such high returns are more likely to be due to effort rather than pure lack, since a monotonic incentive scheme will then motivate the agent to provide more effort. Differentiating (22) one soon finds that the formal requirement is that the likelihood ratio $f_a(x, a)/f(x, a)$ is increasing in $x$, which is known as the monotone likelihood ratio property. Milgrom (1981) demonstrated how the condition could be interpreted to mean that higher returns convey “good news” about the agent’s effort.

**Approximation to the First-Best**

If the likelihood derivative in (23) becomes unbounded as the outcomes tend toward the worst, and the agent’s utility $U(s(x))$ is also unbounded below, incentive compatible contracts can approach arbitrarily close to the first-best by setting $s(x)$ at extremely punishing levels for such outcomes. Of course the exact first-best is not attainable, so technically the problem has no solution. But in practice, one can say that moral hazard can be almost fully overcome under these conditions. Mirrlees [1974b] gave an example of this, and Mirrlees [1975a] proved the general result.

The intuition is clear from the likelihood interpretation. If the likelihood derivative goes to infinity for values of $x$ close to its lower limit, this means that bad outcomes are extremely informative about the agent’s having taken a suboptimal action. When he takes the optimal action, the bad outcomes are very unlikely; therefore the agent need be given only a small offsetting reward in the case of other outcomes to fulfill his participation constraint. But if he takes a suboptimal action, the bad outcomes become much more likely, and the severe penalties associated with these outcomes

give the agent a large incentive to avoid such actions. Thus the principal can induce first-best effort at near-zero cost.

**Validity of the First-Order Approach**

Before Mirrlees, the replacement of the agent’s maximization by its first-order condition was done routinely and without thought. But Mirrlees [1975a, 1982c] showed that this procedure was in fact very problematic. Under standard assumptions, an agent’s problem under moral hazard need not be convex. Hence, the agent’s optimal action may occur at an extreme of the feasible set of actions. There may also be multiple solutions to the first-order conditions, with some of the solutions yielding local minima or stationary points that are neither maxima nor minima. Mirrlees also asked when the first-order approach is valid. He correctly identified two conditions that are of use, the monotone likelihood ratio property, and the convex distribution function property, but his proof was faulty. Further work by Rogerson (1985) and Jewitt (1988) has now completed this aspect of the theory.

**Application to Hierarchies**

Mirrlees [1976a] applied this general framework to study the internal organization of firms, specifically the use of incentive wage schedules and the role of hierarchies. This is an important contribution to the formalization of Coase’s idea that the size of the firm is determined by trading off the costs of internal control against the costs of using the market. A particularly intriguing result in Mirrlees’s paper is that payment schedules at lower levels of the firm’s internal hierarchy should be more like fixed wages for following particular instructions, and at higher levels there should be more profit-sharing incentives. This paper has also inspired a lot of further work.

**Later Developments and Applications**

As with the revelation principle for the case of adverse selection, the theory of incentive design under moral hazard has become a part of everyday economic theory. We select a few examples for special mention. For an exposition with a somewhat different perspective and using somewhat different techniques, see Stiglitz (1983).

The basic idea that outcomes convey probabilistic information about the underlying effort, and that incentive schemes can be based on any variable that is informative in this manner, has very wide applicability. For example, Holmström (1982) uses it to show how, when a principal has several agents...
and the errors in their outcomes are correlated, the optimal incentive scheme for each agent should use the outcomes of all other agents. When only rank order information about individuals' performance is known, contests emerge as optimal mechanisms, with payment according to the order in which individuals perform.

Grossman and Hart (1983) construct a very general model, using an alternative that circumvents the difficulties of the first-order approach. However, the analysis is at such a level of generality that few specific results emerge. On the other hand, Holmström and Milgrom (1987) offer a very specific model based on constant absolute risk aversion and a normal distribution of outcomes that arises from aggregation of numerous small shocks through time. This enables them to obtain and characterize an optimal incentive scheme that is linear in outcomes. Its coefficients depend in a natural way on the agent's risk aversion and disutility of effort, and on the error with which the observed outcome reflects the unobservable effort. If the outcome is a more accurate measure of the effort, then the optimal incentive scheme will be "more powerful" in the sense of giving a higher marginal reward for output.

Analysis of moral hazard in the context discussed by Mirrlees has found many applications. Mirrlees focused at length on the internal workings of the firm. Later work on this theme has considered more sophisticated incentive schemes that include auditing; see, for example, Dye (1986). This has a large number of applications in the accounting literature. The work has also been developed to study optimal financial structure from a principal agent point of view. Formal developments of ideas in Jensen and Meckling's (1976) essay owes a great deal to the ideas developed in Mirrlees' work. In all of these applications, the trade-off between risk sharing and moral hazard is a central concern.

A major drawback with the above formulation is that effort has only one dimension. In real world situations agents typically have many different dimensions of discretion. Output too may be multi-dimensional with different types of output being differentially responsive to particular kinds of effort. Holmström and Milgrom (1987, 1991) initiate a study such multi-task environments. Of particular importance is the interaction between different kinds of activities. Compare two activities, whose outcomes measure their inputs with different degrees of accuracy. Considered in isolation, the more accurately measured activity should have a more powerful incentive scheme. But suppose the same agent performs both, and they are substitutes in the sense that more effort devoted to one comes at the cost of effort devoted to the other. Offering a higher-powered incentive scheme to the more accurately measured activity will divert the agent's effort away from the other. Therefore the principal finds it optimal to weaken the incentives for this activity too. This work is proving to be one

of the most fruitful extensions of the basic moral hazard model; for example, Dixit (1996, appendix) uses it to explain the weakness of incentives in government bureaucracies.

V. Mirrlees’ Other Contributions

While the contributions discussed above go to the very foundations of economic theory, Mirrlees is perhaps even better known to most of the economics profession for his specific contributions to public economics and development economics.

His joint work with Peter Diamond on commodity taxation was a major advance upon previous treatments of Frank Ramsey and Paul Samuelson; the Diamond–Mirrlees model now stands as the definitive statement and analysis of this branch of public economics.

The second of the two Diamond–Mirrlees papers [1971a] and [1971b] elaborates Ramsey–Samuelson type inverse-elasticity tax formulas, and is probably the more widely known. Many subsequent papers have been written, expressing the formulas slightly differently, examining their implications in special contexts like intertemporal allocation or international trade, and making them amenable to empirical work.

But the first Diamond–Mirrlees paper contains a far more fundamental result. Since commodity taxation drives a wedge between consumer and producer prices, it necessarily leaves us in a second-best world. The general theory of such an economy offers little hope for any clear prescriptions; we know from the negative result of Lipsey and Lancaster that once one condition of Pareto optimality is violated, then it may be desirable to violate others. But Diamond and Mirrlees offer a positive result of surprising generality: under a wide range of conditions it remains desirable to keep the production sector as a whole operating efficiently. Thus intermediate inputs should not be taxed or subsidized, a small country should keep its production sector exposed to world prices, and in the intertemporal context interest should be taxed or subsidized, if so desired, only at the stage where consumers or savers receive such income.

The argument is remarkably simple. Make two assumptions: (i) the aggregate demand vector varies continuously with respect to the set of tax instruments, and (ii) the set of instruments is sufficiently rich that a variation that potentially benefits all consumers can be found. Suppose a tax equilibrium leaves the aggregate demand point in the interior of the feasible aggregate production set. Then by (ii) a direction of variation can be found to benefit all consumers, and by (i) a small move in this direction is feasible. Therefore the initial point cannot have been optimal. A uniform lump-sum subsidy satisfies these conditions. More interestingly, so does
commodity taxation, provided there is a commodity (which may be a Hicksian composite) that is not simultaneously positively demanded by one consumer and positively supplied by another. An increase in its price must be Pareto improving if no consumer is a net demander, and a decrease works if none is a net supplier.

In later work they extended this result. For example Mirrlees [1972a] considered how restricted ability to tax profits can justify departures from production efficiency, and Diamond and Mirrlees [1976b] built on earlier work by both authors to show that under constant returns to scale, private production should continue to make zero profits at the correct shadow prices. This formulation provided the most general theoretical underpinning for Mirrlees’ work on cost benefit analysis with Ian Little.

Little and Mirrlees applied these ideas to derive very detailed criteria and procedures for evaluating development projects. The production efficiency result has important policy implications, especially for less developed countries where production distortions abound. Their manual [1974c] had significant impact on the thinking and practices of international organizations, particularly the World Bank, which evaluate and support such projects.

VI. Concluding Remarks

Mirrlees’ work has enhanced our understanding of a range of important problems. While setting out to study specific applications, mostly in public economics and development economics, he has developed analytic insights that had beneficial consequences far beyond the problem at hand. Thus, not only has he solved important problems, he has provided us with a novel way to address different problems in the future.

The fact that Mirrlees began with applications, and not with the expressed aim of developing general theories is an important aspect of his approach. Thus, his attempts to put the welfare economics of government policy on a rigorous footing stem from an intrinsic interest in providing policy guidance. Hence, while one learns a great deal about economic analysis in general from studying his contributions, one is also reminded of his vision that one reason for doing economics is the possibility of making the world a better place.

References


Myerson, R. B.: Optimal auction design. Mathematics of Operations Research 6, 58–73,


