~~a given product, then even if their proportions are variable, there is still no principle by which the market can determine their relative prices except by arbitrary bargaining (Mises, 1949, p. 336). In the real world, of course, the existence of such purely specific factors, and hence the scope for such bargaining, will be extremely limited.~~

~~The other important point is that values cannot be added or divided, and that the imputation process takes place, not automatically or precisely in an abstract realm of 'values', but only concretely and by trial and error, in the realistic market process of changing prices. In other words, although consumers can evaluate consumer goods and determine their prices directly by valuation, the prices of productive factors are only determined indirectly through market prices and entrepreneurial trial and error. There is no direct, abstract or pure process of imputing values.~~

~~This problem became strikingly relevant during the well-known debate over the Mises–Hayek demonstration that socialist governments cannot calculate economically. Joseph Schumpeter brusquely dismissed this contention with the statement that economic calculation under socialism follows 'from the elementary proposition that consumers in evaluating ("demanding") consumers' goods *ipso facto* also evaluate the means of production which enter into the production of these goods' (Schumpeter, 1942, p. 175). Hayek's perceptive reply points out that the '*ipso facto*' assumes complete knowledge of values, demands, scarcities, etc. to be 'given' to everyone, thereby ignoring the reality of the universal lack of complete knowledge, as well as the necessary function of the market economy, and the market price system, in conveying knowledge to all its participants (Hayek, 1945).~~

~~The analysis of imputation began in a neglected work of Aristotle, the *Topics*. Here, Aristotle analysed the ends–means relationship, and pointed out that the means, or 'instruments of production', necessarily derive their value from the ends, the final products useful to man, 'the instruments of action'. The more desirable the final good, the more valuable will be the means to arrive at the product. Aristotle introduced the theme of marginality by stating that if the addition of a good *A* to an already desirable good *C* yields a more desirable result than the addition of good *B*, then *A* will be more highly valued than *B*. Indeed, he also added a pre-Böhm-Bawerkian note by stressing the differential value of the loss rather than the addition of a good. Good *A* will be more valuable than *B*, if the loss of *A* is considered to be worse than the loss of *B*. While critics have noted that Aristotle only slightly applied his analysis to the economic realm, his imputation theory was still an important contribution to the general theory of action of which economic theory is a highly developed part (Spengler, 1955).~~

~~MURRAY N. ROTHBARD~~

~~*See also* ADDING-UP PROBLEM; AUSTRIAN SCHOOL OF ECONOMICS; MARGINAL PRODUCTIVITY THEORY.~~

~~BIBLIOGRAPHY~~

~~Aristotle. *Topica*. Trans. into English by W.A. Pickard-Cambridge, and included in Vol. I of *The Works of Aristotle*, ed. W.D. Ross, Oxford: Clarendon Press, 1928.~~

~~Hayek, F.A. 1926. Some remarks on the problem of imputation. In F.A. Hayek, *Money, Capital, and Fluctuations: early Essays*, Chicago: University of Chicago Press, 1984.~~

~~Hayek, F.A. 1945. The use of knowledge in society. In F.A. Hayek, *Individualism and Economic Order*, Chicago: University of Chicago Press, 1948.~~

~~Kauder, E. 1965. *A History of Marginal Utility Theory*. Princeton: Princeton University Press.~~

~~Menger, C. 1871. *Principles of Economics*. Glencoe, Ill.: Free Press, 1950.~~

~~Mises, L. von. 1949. *Human Action: a treatise on economics*. 3rd edn, Chicago: Regnery, 1966.~~

~~Schumpeter, J.A. 1942. *Capitalism, Socialism, and Democracy*. New York: Harper.~~

~~Spengler, J. 1955. Aristotle on economic imputation and related matters. *Southern Economic Journal* 21, April, 371–89.~~

~~Stigler, G. 1941. *Production and Distribution Theories: The Formative Period*. New York: Macmillan.~~

~~Wieser, F. von. 1889. *Natural Value*. New York: Kelley & Millman, 1956.~~

incentive compatibility. Allocation mechanisms, organizations, voting procedures, regulatory bodies, and many other institutions are designed to accomplish certain ends such as the Pareto-efficient allocation of resources or the equitable resolution of disputes. In many situations it is relatively easy to conceive of feasible processes; processes which will accomplish the goals if all participants follow the rules and are capable of handling the informational requirements. Examples of such mechanisms include marginal cost pricing, designed to attain efficiency, and equal division, designed to attain equity. Of course once a feasible mechanism is found, the important question then becomes whether such a mechanism is also informationally feasible and compatible with 'natural' incentives of the participants. Incentive compatibility is the concept introduced by Hurwicz (1972, p. 320) to characterize those mechanisms for which participants in the process would not find it advantageous to violate the rules of the process.

The historical roots of the idea of incentive compatibility are many and deep. As was pointed out in one of a number of recent surveys,

> the concept of incentive compatibility may be traced to the 'invisible hand' of Adam Smith who claimed that in following individual self-interest the interests of society might be served. Related issues were a central concern in the 'Socialist Controversy' which arose over the viability of a decentralized socialist society. It was argued by some that such societies would have to rely on individuals to follow the rules of the system. Some believed this reliance was naive; others did not. (Groves and Ledyard, 1986, p. 1).

Further, the same issues have arisen in the design of voting procedures. Concepts and problems related to incentives were already identified and documented in the 18th century in discussions of proposals by Borda to provide alternatives to majority rule committee decisions. (See STRATEGY-PROOF ALLOCATION MECHANISMS for further information on voting procedures.)

Incentive compatibility is both desirable and elusive. The desirability of incentive compatibility can be easily illustrated by considering public goods, goods such that one consumer's consumption of them does not detract from another consumer's simultaneous consumption of that good. The existence of these collective consumption commodities creates a classic situation of *market failure*; the inability of markets to arrive at a Pareto-optimal allocation. It was commonly believed, prior to Groves and Ledyard (1977), that in economies with public goods it would be impossible to devise a decentralized process that would allocate resources efficiently since agents would have an incentive to 'free ride' on others' provision of those goods in order to reduce their own share of providing them. Of course Lindahl (1919) had proposed a feasible process which mimicked markets by creating a separate price for each individual's consumption of the public good. This designed process was, however, rejected as unrealistic by those who recognized that these 'synthetic markets' would be shallow (essentially monopsonistic) and therefore buyers would have no incentive to treat prices as fixed and invariant to their demands. The classic quotation is '... it is in the selfish interest of each person to give *false* signals, to pretend to have less interest in a given collective consumption activity than he really has...' (Samuelson, 1954, pp. 388–9). Allocating public goods efficiently through Lindahl pricing would be feasible and successful if consumers followed the rules; but, it would not be successful since the mechanism is not incentive compatible. If buyers do not follow the rules, efficient resource allocation will not be achieved and the goals of the design will be subverted because of the motivations of the participants. Any institution or rule, designed to accomplish group goals, must be incentive compatible if it is to perform as desired.

The elusiveness of incentive compatibility can be most easily illustrated by considering a situation with only private goods. Economists generally model behaviour in private goods markets by assuming that buyers and sellers 'follow the rules' and take prices as given. It is now known, however, that as long as the number of agents is finite then any one of them can still gain by misbehaving and, furthermore, can do so in a way which can not be detected by anyone else. The explanation is provided in two steps. First, if there are a finite number of traders, and none have a perfectly elastic offer curve (which will be true if preferences are non-linear) then one trader can gain by being able to control prices. For example, a buyer would want to set price where his marginal benefit equalled his marginal outlay and thereby gain monopsonistic benefits. Of course, if the others know that buyer's demand curve (either directly or through inferences based on revealed preference) then they would know that the buyer was not 'taking prices as given' and could respond with a suitable punishment against him. This brings us to our second step. Even though others can monitor and prohibit price setting behaviour, our benefit-seeking monopsonist has another strategy which can circumvent this supervision. He calculates a (false) demand curve which, when added to the others' offer curves, produces an equilibrium price equal to that which he would have set if he had direct control. He then calculates a set of preferences which yields that demand curve and participates in the process *as if he had these (false) preferences.* Usually this involves simply acting as if one has a slightly lower demand curve than one really does. Since preferences are not able to be observed by others, he can follow this behaviour which looks like it is price-taking, and therefore 'legal', and can do individually better. The unfortunate implication of such concealed misbehaviour is that the mechanism performs other than as intended. In this case, resources are artificially limited and too little is traded to attain efficiency.

In 1972 Hurwicz established the validity of the above intuition. His theorem can be precisely stated after the introduction of some notation and a framework for further discussion.

THE IMPOSSIBILITY THEOREM. The key concepts include economic environments, allocation mechanisms, incentive compatibility, the no-trade option, and Pareto-efficiency. We take up each in turn.

An *economic environment*, those features of an economy which are to be taken as given throughout the analysis, includes a description of the agents, the feasible allocations they have available and their preferences for those allocations. While many variations are possible, I concentrate here on a simple model. Agents (consumers, producers, politicians, etc.) are indexed by $i = 1, \ldots, n$. X is the set of feasible allocations where $x = (x^i, \ldots, x^n)$ is a typical element of X. (An exchange environment is one in which X is the set of all $x = (x^1, \ldots, x^n)$ such that $x^i \geqslant 0$ and $\Sigma x^i = \Sigma w^i$, where w^i is i's initial endowment of commodities.) Each agent has a selfish utility function $u^i(x^i)$. The environment is $e = [I, X, u^1, \ldots, u^n]$. A crucial fact is that initially *information is dispersed* since i, and only i, knows u^i. We identify the specific knowledge i initially has as i's *characteristic*, e^i. In our model, $e^i = u^i$.

Although there are many variations in models of allocation mechanisms, I begin with the one introduced by Hurwicz (1960). An *allocation mechanism* requests information from the agents and then computes a feasible allocation. It requests information in the form of messages m^i from agent i through a *response function* $f^i(m^i, \ldots, m^n)$. Agent i is told to report $f^i(m, e^i)$ if others have reported m and i's characteristic is e^i. An equilibrium of these response rules, for the environment e, is a joint message m such that $m^i = f^i(m, e^i)$ for all i. Let $\mu(\mathrm{e}, f)$ be the set of equilibrium messages for the response functions f in the environment e. The allocation mechanism computes a feasible allocation x by using an *outcome function* $g(m)$ on equilibrium messages. The net result of all of this in the environment e is the allocation $g[\mu(e, f)] = x$ *if all i follow the rules, f.* Thus, for example, the *competitive mechanism* requests agents to send their demands as a function of prices which are in turn computed on the basis of the aggregate demands reported by the consumers. In equilibrium, each agent is simply allocated their stated demand. (An alternative mechanism, yielding exactly the same allocation in one iteration, would request the demand *function* and then compute the equilibrium price and allocation for the reported demand functions.) It is well known, for exchange economies with only private goods, that if agents report their true demands then the allocations computed by the competitive mechanism will be Pareto-optimal.

It is obviously important to be able to identify those mechanisms, those rules of communication, that have the property that they are self-enforcing. We do that by focusing on a class of mechanisms in which each agent gains nothing, and perhaps even loses, by misbehaving. While a multitude of misbehaviours could be considered it is sufficient for our purposes to consider a slightly restricted range. In particular we can concentrate on undetectable behaviour, behaviour which no outside agent can distinguish from that prescribed by the mechanism. We model this limitation on behaviour by requiring the agent to restrict his misrepresentations to those which are consistent with some characteristic he might have. An allocation mechanism is said to be *incentive compatible* for all environments in the class E if there is no agent i and no environment e in E and no characteristic e^{*i} such that (e/e^{*i}) is in E (where (e/e^{*i}) is the environment derived from e by replacing e^i with e^{*i}) and such that

$$u^i\{g[\mu(e, f)], e^i\} > u^i\{g[\mu(e/e^{*i}, f], e^i\}$$

where $u^i(x^*, e^i)$ is i's utility function in the environment e. That is, no agent can manipulate the mechanism by pretending to have a characteristic different from the true one and do better than acting according to the truth. The agent has an incentive to follow the rules and the rules are compatible with his motivations.

Incentive compatibility is at the foundation of the modern *theory of implementation*. In that theory, one tries to identify conditions under which a particular social choice rule or performance standard, $P : E \rightarrow X$, can be recreated by an allocation mechanism under the hypothesis that individuals will follow their self-interest when they participate in the implementation process. In our language, the rule P is implementable if and only if there is an incentive compatible mechanism (f, g) such that $g[\mu(e,f)] = P(e)$ for all e in E. The theory of implementation seeks to answer the question 'which P are implementable?' We will see some of the answers below for P which select from the set of Pareto-efficient allocations. Those interested in more general goals and performance standards should consult Dasgupta, Hammond and Maskin (1979) or Postlewaite and Schmeidler (1986).

An allocation mechanism is said to have the *no trade-option* if there is an allocation θ at which each participant may remain. In exchange environments the initial endowment is usually such an allocation. Mechanisms with a no-trade option are non-coercive in a limited sense. If an allocation mechanism possesses the no-trade option then the allocation it computes for an environment e, if agents follow the rules, must leave everyone at least as well off, using the utility functions for e, as they are at θ. That is, for all i and all e in E

$$u^i\{g[\mu(e,f)], e^i\} > u^i(\theta, e^i).$$

An allocation mechanism is said to be *Pareto-efficient* in E if the allocations selected by the mechanism, when agents follow the rules, are Pareto-optimal in e. That is, for each e in E there is no allocation x^* in X such that, for all i,

$$u^i(x^*, e^i) \geqslant u^i\{g[\mu(e,f)], e^i\}$$

with strict inequality for some i.

With this language and notation, Hurwicz's theorem on the elusive nature of incentive compatibility in private markets, subsequently expanded by Ledyard and Roberts (1974) to include public goods environments, can now be easily stated. *Theorem*: In classical (public or private) economic environments with a finite number of agents, there is no incentive compatible allocation mechanism which possesses the no-trade option and is Pareto-efficient. (Classical environments include pure exchange environments with Cobb–Douglas utility functions.)

A more general version of this theorem, in the context of social choice theory, has been proven by Gibbard (1973) and Satterthwaite (1975) with the concept of a 'non-dictatorial social choice function' replacing that of a 'mechanism with the no-trade option'. (See STRATEGY-PROOF ALLOCATION MECHANISMS.)

There are a variety of possible reactions to this theorem. One is simply to give up the search for solutions to market failure since the theorem seems to imply that one should not waste any effort trying to create institutions to allocate resources efficiently. A second is to notice that, at least in private markets, if there are a very large number of individuals in each market then efficiency is 'almost' attainable (see Roberts and Postlewaite, 1976). A third is to recognize that the behaviour of individuals will generally be different from that implicitly assumed in the definition of incentive compatibility. A fourth is to accept the inevitable, lower one's sights, and look for the 'most efficient' mechanism among those which are incentive compatible and satisfy a voluntary participation constraint. We consider the last two options in more detail.

OTHER BEHAVIOUR: NASH EQUILIBRIUM. If a mechanism is incentive compatible, then each agent knows that his best strategy is to follow the rules according to his true characteristic, *no matter what the other agents will do.* Such a strategic structure is referred to as a dominant strategy game and has the property that no agent need know or predict anything about the others' behaviour. In mechanisms which are not incentive compatible, each agent must predict what others are going to do in order to decide what is best. In this situation agents' behaviour will not be as assumed in the definition of incentive compatibility. What it will be continues to be an active research topic and many models have been proposed. Since most of these are covered in Groves and Ledyard (1986), I will concentrate on the two which seem most sensible. Both rely on game-theoretic analyses of the strategic possibilities. The first concentrates on the outcome rule, g, and postulates that agents will not choose messages to follow the specifications of the response functions but to do the best they can against the messages sent by others. Implicitly this assumes that there is some type of iterative process (embodied in the response rules) which allows revision of one's message in light of the responses of others. We can formalize this presumed strategic behaviour in a new concept of incentive compatibility. An allocation mechanism (f, g) is called *Nash incentive compatible* for all environments in E if there is no environment e, no agent i, and no message m^{*i} which i can send such that

$$u^i(g[\mu(e,f)/m^{*i}, e^i]) > u^i(g[\mu(e,f), e^i])$$

where $\mu(e,f)$ is the 'equilibrium' message of the response rules f in the environment e, $g(m)$ is the outcome rule, and $[m/m^{*i}]$ is the vector m where m^{*i} replaces m^i. In effect this requires the equilibrium messages of the response rules to be Nash equilibria in the game in which messages are strategies and payoffs are given by $u[g(m)]$. It was shown in a sequence of papers written in the late 1970s, including those by Groves and Ledyard (1977), Hurwicz (1979), Schmeidler (1980), and Walker (1981), that Nash incentive compatibility is not elusive. The effective output of that work was to establish the following. *Theorem*: In classical (public or private) economic environments with a finite number of agents, there are many Nash incentive compatible mechanisms which possess the no-trade option and are Pareto-efficient.

With a change in the predicted behaviour of the participants in the mechanism, in recognition of the fact that in the absence of dominant strategies agents must follow some other self-interested strategies, the pessimism of the Hurwicz theorem is replaced by the optimistic prediction of a plethora of possibilities. (See Dasgupta, Hammond and Maskin, 1979, Postlewaite and Schmeidler (1986) and Groves and Ledyard (1986) for comprehensive surveys of these results including many for more general social choice environments.) Although it remains an unsettled empirical question whether participants will indeed behave this way, there is a growing body of experimental evidence that seems to me to support the behavioural hypotheses underpinning Nash incentive compatibility, especially in iterative tâtonnement processes.

OTHER BEHAVIOUR: BAYES' EQUILIBRIUM. The second approach to modelling strategic behaviour of agents in mechanisms, when dominant strategies are not available, is based on Bayesian decision theory. These models, called *games of incomplete information* (see Myerson, 1985), concentrate on the beliefs of the players about the situation in which they find themselves. In the simplest form, it is postulated that there is a common knowledge (everyone knows that everyone knows that ...) probability function, $\pi(e)$, which describes everyone's prior beliefs. Each agent is then assumed to choose that message which is best against the expected behaviour of the other agents. The expected behaviour of the other agents is also constrained to be 'rational' in the sense that it should be best against the behaviour of others. This presumed strategic behaviour is embodied in a third type of incentive compatibility. (It could be argued that the concept of incentive compatibility remains the same, based on non-cooperative behaviour in the game induced by the mechanism, while only the presumed information structure and sequence of moves required to implement the allocation mechanism are changed. Such a view is not inconsistent with that which follows.) An allocation mechanism (f, g) is called *Bayes incentive compatible* for all environments in E given π on E if there is no environment e^*, no agent i, and no message m^{*i} which i can send such that

$$\int u^i\{g[\mu(e, f)/m^{*i}], e^{*i}\}\,\mathrm{d}\pi(e\,|\,e^{*i})$$
$$> \int u^i\{g[\mu(e, f), e^{*i}]\,\mathrm{d}\pi(e,|\,e^{*i})\}$$

where, as before, μ is the equilibrium message vector and g is the outcome rule. Further, $\pi(e\,|\,e^{*i})$ is the conditional probability measure on e given e^{*i}, and u^i is a von Neumann–Morgenstern utility function. In effect, this requires the equilibrium messages of the response rules to be Bayes equilibrium outcomes of the incomplete information game with messages as strategies, payoffs $u[g(m)]$ and common knowledge prior π.

There are two types of results which deal with the possibilities for Bayes incentive compatible design of allocation mechanisms, neither of which is particularly encouraging. The first type deals with the possibilities for incentive compatible design which is independent of the beliefs. The typical theorem is illustrated by the following result proven by Ledyard (1978). *Theorem*: In classical economic environments with a finite number of agents, there is no Bayes incentive compatible mechanism which possesses the no-trade option and is Pareto-efficient *for all π on E*. Understanding this result is easy when one realizes that any mechanism (f, g) is Bayes incentive compatible for all π for all e in E if and only if it is (Hurwicz) incentive compatible for all e in E. Thus the Hurwicz impossibility theorem again applies.

The second type of result is directed towards the possibilities for a specific prior π; that is, towards what can be done if the mechanism can depend on the common knowledge beliefs. The most general characterizations of the possibilities for Bayes incentive compatible design can be found in Palfrey and Srivastava (1987) and Postlewaite and Schmeidler (1986). They have shown that two conditions, called monotonicity and self-selection, are necessary and sufficient for a social choice correspondence to be implementable in the sense that there is a Bayes incentive compatible mechanism that reproduces that correspondence. The details of these conditions are not important. What is important is that many correspondences do not satisfy them. In particular, there appear to be many priors π and many sets of environments E for which there is no mechanism which is Bayes incentive compatible, provides a no-trade option and is Pareto-efficient. Thus, impossibility still usually occurs even if one allows the mechanism to depend on the prior.

One recent avenue of research which promises some optimistic counterweight to these negative results can be found in Palfrey and Srivastava (1987). In much the same way that the natural move from Hurwicz incentive compatibility to Nash incentive compatibility created opportunities for incentive compatible design, these authors have shown that a move back towards dominant strategies may also open up possibilities. Refinements arise by varying the equilibrium concept in a way that reduces the number of (Bayes or Nash) equilibria for a given e or π. Moore and Repullo use subgame perfect Nash equilibria. Palfrey and Srivastava eliminate weakly dominated strategies from the set of Nash equilibria. They have discovered that, in pure exchange environments, virtually all performance correspondences are implementable if behaviour satisfies these refinements. In particular, any selection from the Pareto-correspondence is implementable for these refinements, and so there are many refined-Nash incentive compatible mechanisms which are Pareto-efficient and allow a no-trade option. It is believed that these results will transfer naturally to refinements of Bayes equilibria, but the research remains to be done.

INCENTIVE COMPATIBILITY AS A CONSTRAINT. Another of the reactions to the Hurwicz impossibility result is to accept the inevitable, to view incentive compatibility as a constraint, and to design mechanisms to attain the best level of efficiency one can. If full efficiency is possible, it will occur as the solution. If not, then one will at least find the second-best allocation mechanism. Examples of this rapidly expanding research literature include work on optimal auctions (Harris and Raviv, 1981; Matthews, 1983; Myerson, 1981), the design of optimal contracts for the principle-agent problem, and the theory of optimal regulation (Baron and Myerson, 1982). As originally posed by Hurwicz (1972, pp. 299–301), the idea is to adopt a social welfare function $W(x, e)$, a measure of the social welfare attained from the allocation x if the environment is e and then to choose the mechanism (f, g) to maximize the (expected) value of W subject to the 'incentive compatibility constraints', the constraint that the rules (f, g) be consistent with the motivations of the participants. One chooses (f, g) to

$$\text{maximize} \int W\{g[\mu(e, f)], e\}\,\mathrm{d}\pi(e)$$

subject to, for every i, every e, and every e^{*i},

$$\int u^i\{g[\mu(e/e^{*i}, f)], e^i\}\,\mathrm{d}\pi(e\,|\,e^i) \leqslant \int u^i(g[\mu(e, f), e^i])\,\mathrm{d}\pi(e\,|\,e^i).$$

As formalized here the incentive compatibility constraints embody the concept of Bayes incentive compatibility. Of course, other behavioural models could be substitued as appropriate.

Sometimes a voluntary participation constraint, related to the no-trade option of Hurwicz, is added to the optimal design problem. One form of this constraint requires that (f, g) also satisfy, for every i and every e,

$$\int u^i\{g[\mu(e)], e^i\}\,\mathrm{d}\pi(e\,|\,e^i) \geqslant \int u^i(\theta[e], e^i)\,\mathrm{d}\pi(e\,|\,e^i).$$

In practice this optimization can be a difficult problem since there are a large number of possible mechanisms (f, g). However, an insight due to Gibbard (1973) can be employed to reduce the range of alternatives and simplify the analysis. Now called the *revelation principle*, the observation he made was that, to find the maximum, it is sufficient to consider only mechanisms, called direct revelation mechanisms, in which agents are asked to report their own characteristics. The reason is easy to see. Suppose that (f^*, g^*) solves the maximum problem. Let (F^*, G^*) be a new (direct revelation) mechanism defined by $F^{*i}(m, e^i) = e^i$ and $G^*(m) = g[\mu(m, f)]$. Each i is told to report his characteristic and then G^* computes the allocation by computing that which would have been chosen if the original mechanism (f, G^*) had been used honestly in the reported environment. (F^*, G^*) yields the same allocation as (f^*, g^*), *if each agent reports the truth*. But the incentive compatibility constraints, which (f^*, g^*) satisfied, ensure that each agent will want to report truthfully. Thus, whatever can be done, by any arbitrary mechanism subject to the Bayes incentive compatibility constraints, can be done with direct revelation mechanisms subject to the constraint that each agent wants to report their true characteristic. One need only choose a function $G: E \rightarrow X$ to

$$\text{maximize} \int W[G(e), e]\, \mathrm{d}\pi(e)$$

subject to, for every i, e and e^i,

$$\int u^i[G(e/e^{*i}), e^i]\, \mathrm{d}\pi(e\,|\,e^i) \leqslant \int u^i[G(e), e^i]\, \mathrm{d}\pi(e\,|\,e^i),$$

and

$$\int u^i[G(e), e^i]\, \mathrm{d}\pi(e\,|\,e) \geqslant \int u^i(\theta[e], e^i)\, \mathrm{d}\pi(e\,|\,e^i).$$

There are at least two problems with this approach to organizational design. The first is that the choice of mechanism depends crucially on the prior beliefs, π. This is a direct result of the use of Bayes incentive compatibility in the constraints. Since the debate is still open let me simply summarize some of the arguments. One is that if the mechanism chosen for a given situation does not depend on common knowledge beliefs then we would not be using all the information at our disposal to pursue the desired goals and would do less than is possible. Further, since the beliefs are common knowledge we can all agree as to their validity (misrepresentation is not an issue) and therefore to their legitimate inclusion in the calculations. An argument is made against this on the practical grounds that one need only consider actual situations, such as the introduction of new technology by a regulated utility or the acquisition of a major new weapons system by the government, to understand the difficulties involved in arriving at agreements about the particulars of common knowledge. Another argument against is based on the feeling that mechanisms should be robust. A 'good' mechanism should be able to be described in terms of its mechanics and, while it probably should have the capacity to incorporate the common knowledge relevant to the current situation, it should be capable of being used in many situations. How to capture these criteria in the constraints or the objective function of the designer remains an open research question.

The second problem with the optimal auction approach to organizational design is the reliance on the revelation principle. Restricting attention to direct revelation mechanisms, in which an agent reports his entire characteristic, is an efficient way to prove theorems, but it provides little guidance for those interested in actual organization design. For example it completely ignores the informational requirements of the process and any limitations, if any, in the information processing capabilities of the agents or the mechanism. Writing down one's preferences for all possible consumption patterns is probably harder than writing down one's entire demand surface which is certainly harder than simply reacting to a single price vector and reporting only the quantities demanded at that price. A failure to recognize the information processing constraints in the optimization problem is undoubtedly one of the reasons there has been limited success in using the theory of optimal auctions to explain the existence of pervasive institutions, such as the first-price sealed-bid auction used in competitive contracting or the posted price institution used in retailing.

SUMMARY. Incentive compatibility captures the fundamental positivist notion of self-interested behaviour that underlies almost all economic theory and application. It has proven to be an organizing principle of great scope and power. Combined with the modern theory of mechanism design, it provides a framework in which to analyse such diverse topics as auctions, central planning, regulation of monopoly, transfer pricing, capital budgeting, and public enterprise management. Incentive compatibility provides a basic constraint on the possibilities for normative analysis. As such it serves as the fundamental interface between what is desirable and what is possible in a theory of organizations.

JOHN O. LEDYARD

See also BIDDING; EFFICIENT ALLOCATION; EXTERNALITY; LINDAHL EQUILIBRIUM; ORGANIZATION THEORY; PUBLIC GOODS; REVELATION OF PREFERENCES.

BIBLIOGRAPHY

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

Dasgupta, P., Hammond, P. and Maskin, E. 1979. The implementation of social choice rules: some general results on incentive compatibility. *Review of Economic Studies* 46, 185–216.

Gibbard, A. 1973. Manipulation of voting schemes: a general result. *Econometrica* 41, 587–602.

Groves, T. and Ledyard, J. 1977. Optimal allocation of public goods: a solution to the 'free rider' problem. *Econometrica* 45, 783–809.

Groves, T. and Ledyard, J. 1986. Incentive compatibility ten years later. In *Information, Incentives, and Economic Mechanisms*, ed. T. Groves, R. Radner and S. Reiter. Minneapolis: University of Minnesota Press.

Harris, M. and Raviv, A. 1981. Allocation mechanisms and the design of auctions. *Econometrica* 49, 1477–99.

Hurwicz, L. 1960. Optimality and informational efficiency in resource allocation processes. In *Mathematical Methods in the Social Sciences*, ed. K. Arrow, S. Karlin and P. Suppes, Stanford: Stanford University Press, 27–46.

Hurwicz, L. 1972. On informationally decentralized systems. In *Decision and Organization: A Volume in Honor of Jacob Marschak*, ed. R. Radner and C.B. McGuire, Amsterdam: North-Holland, 297–336.

Hurwicz, L. 1979. Outcome functions yielding Walrasian and Lindahl allocations at Nash equilibrium points. *Review of Economic Studies* 46, 217–25.

Ledyard, J. 1978. Incomplete information and incentive compatibility. *Journal of Economic Theory* 18, 171–89.

Ledyard, J. and Roberts, J. 1974. On the incentive problem with public goods. Discussion Paper No. 116, Center for Mathematical Studies in Economics and Management Science, Northwestern University.

Lindahl, E. 1919. *Die Gerechtigkeit der Besteuerung*. Lund. Partial translation in *Classics in the Theory of Public finance*, ed. R.A. Musgrave and A.T. Peacock, London: Macmillan, 1958.

Matthews, S. 1983. Selling to risk averse buyers with unobservable tastes. *Journal of Economic Theory* 30, 370–400.

Moore, J. and Repullo, R. 1986. Subgame perfect implementation. London School of Economics, Working Paper.

Myerson, R.B. 1981. Optimal auction design. *Mathematics of Operations Research* 6, 58–73.

Myerson, R.B. 1985. Bayesian equilibrium and incentive compatibility: an introduction. In *Social Goals and Social Organization: Essays in Memory of Elisha Pazner*, ed. L. Hurwicz, D. Schmeidler and H. Sonnenschein, Cambridge: Cambridge University Press.

Palfrey, T. and Srivastava, S. 1986. Implementation in exchange economies using refinements of Nash equilibrium. Graduate School of Industrial Administration, Carnegie-Mellon University, July 1986.

Palfrey, T. and Srivastava, S. 1987. On Bayesian implementable allocations. *Review of Economic Studies* 54(2), 193–208.

Postlewaite, A. and Schmeidler, D. 1986. Implementation in differential information economics. *Journal of Economic Theory* 39(1), June, 14–33.

Roberts, J. and Postlewaite, A. 1976. The incentives for price-taking behavior in large economies. *Econometrica* 44, 115–28.

Samuelson, P. 1954. The pure theory of public expenditure. *Review of Economics and Statistics* 36, 387–9.

Satterthwaite, M. 1975. Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory* 10, 187–217.

Schmeidler, D. 1980. Walrasian analysis via strategic outcome functions. *Econometrica* 48, 1585–93.

Walker, M. 1981. A simple incentive compatible scheme for attaining Lindahl allocations. *Econometrica* 49, 65–71.

incentive contracts. Incentives are the essence of economics. The most basic concept, demand, considers how to induce a consumer to buy more of a particular good; that is, how to give him an incentive to purchase. Similarly, supply relationships are descriptions of how agents respond with more output or labour to additional compensation.

Incentive contracts arise because individuals love leisure. In order to induce them to forgo some leisure, or put alternatively, to put forth effort, some form of compensation must be offered. The theme of this essay is that different forms of incentive contracts deal with some aspects of the problems better than others. The strength of one type of contract is the weakness of another. The labour market trades off these strengths and weaknesses and thereby selects a set of institutions. In what follows, the development of the literature on incentive contracts is briefly discussed. The emphasis is on concepts rather than specific papers or authors, so the bibliography is far from exhaustive.

To discuss incentive contracts, the most general concepts must be narrowed. This essay does that in two ways. First, attention here is restricted to the labour market. At a more general level, incentive contracts can relate to other areas as well. For example, the government may want to have a space satellite built at the lowest possible cost. To do so, incentives must be set appropriately or the producer may charge too much or fail to meet desired quality standards. This problem is analogous to those that arise in the labour context, but for the most part they are ignored, except when isomorphic with the labour market paradigm. Similarly, the law and economics literature is another area where incentive problems are studied, usually in the context of accident liability (see, for example, Green, 1976; Polinsky, 1980; Shavell, 1980). These specific questions are ignored as well, except as they border on the labour market context. Second, the focus is on observability problems. Standard labour supply functions, where hours of work can be observed and paid, are incentive contracts. However, standard labour supply issues are eliminated from consideration since they are dealt with in other essays in *The New Palgrave*.

GENERAL FRAMEWORK

An employer in a competitive environment must induce a worker to perform at the efficient level of effort or face extinction. The reason is simple: if one employer can, through clever use of an incentive contract, get a worker to perform at a more efficient level, that firm's cost will be lower. Lower costs imply that higher wages can be paid to workers and all workers will be stolen from inefficient firms. As a result, the objective function that is taken as standard for the firm is:

$$\operatorname*{Max}_{F} F(Q, E) - C(E), \tag{1}$$

where Q is output and E is worker effort. Thus $F(Q, E)$ is the compensation schedule that the firm announces to the worker; $C(E)$ is the worker's cost of effort function, to be thought of as the dollar cost associated with supplying effort level E.

The competitive nature of the firm in factor and product markets implies that the firm must maximize worker net wealth as in (1) subject to the zero profit constraint:

$$Q = F(Q, E). \tag{2}$$

Output is defined so that each unit sells for \$1 (the numeraire). Thus (2) merely says that output, Q, must be paid entirely to the worker otherwise another firm could steal the worker away by paying more.

The incentive problem arises because the worker takes the compensation scheme $F(Q, E)$ as given and chooses effort to maximize expected utility. Once the worker has accepted the job, his problem is:

$$\operatorname*{Max}_{E} F(Q, E) - C(E). \tag{3}$$

The worker's effort supply function comes from solving the first-order condition associated with (3) or

$$C'(E) = \frac{\partial F}{\partial Q} \cdot \frac{\partial Q}{\partial E} + \frac{\partial F}{\partial E}, \tag{4}$$

which says that the worker sets the marginal cost of effort equal to its marginal return to him. The transformation of effort into output, (i.e. $\partial Q/\partial E$) depends on the production function. A convenient specification is

$$Q = E + v, \tag{5}$$

so that output is the sum of effort, E, and luck, v.

An incentive contract selects $F(Q, E)$ subject to the zero-profit constraint, (2), taking into account that the worker behaves according to (4). There are an infinite variety of incentive contracts that are subsumed by $F(Q, E)$. To make things clear, we consider two polar extremes – the salary and the piece rate (for a more detailed treatment, see Lazear, 1986).

Let us define a salary as compensation that depends only on input so that $F(Q, E)$ takes the form $S(E)$. An hourly wage is an example. Irrespective of the amount that is produced during the hour, the worker receives a fixed amount that depends only on the fact that he supplies E of effort for the hour. (Of course, difficulty in measuring E may be a compelling reason to avoid this form of incentive contract.) At the other extreme is a piece rate where compensation depends only on output so that $F(Q, E)$ takes the form of $R(Q)$. There, no matter how much or how little effort the worker exerts, his compensation depends only on the number of units produced. Both salaries and piece rates are incentive contracts; the first provides incentives by paying workers on the basis of input. The second provides incentives by paying on the basis of output. More sophisticated incentive contracts, which blend the two or use multiperiod approaches are discussed later.

THE PRINCIPAL–AGENT PROBLEM

At the centre of the incentive contract literature is the 'principal–agent' problem. The principal, say, an employer, wants to induce its agent, say, a worker, to behave in a way that is beneficial to the employer. The problem is that the principal's knowledge is imperfect; either he cannot see what the agent does (as in the case of a taxi driver who can sleep on the job) or he cannot interpret the actions (as in the case of an auto mechanic who replaces a number of parts to correct a perhaps simple malfunction). The incentive contracts that can