Introduction

It is surprising to observe that Schumpeter (1954) does not mention the word of incentives in his monumental history of economic thought. How is it possible when today, for many economists, economics is to a large extent a matter of incentives: Incentives to work hard, incentive to produce good quality products, incentives to study, incentives to invest, incentives to save,... How to design institutions in order to provide good incentives for economic agents is a central question of economics today.

Maybe, it is because economics has mostly concentrated on understanding the theory of value in large economies. No eclassical economics in particular postulates rational individual behavior in the market. In a perfectly competitive market, this translates for firms' owners into profit maximization which implies cost minimization. In order words, the pressure of competitive markets solves the problem of incentives for cost minimization. Similarly, consumers faced with exogenous prices have the proper incentives for maximizing their utility levels. The major project of understanding how prices are formed in competitive markets can proceed without worrying about incentives.

However, by treating the firm as a black box, the theory remains silent on how the owners of firms succeed in aligning the objectives of its various members like workers, supervisors, managers with profit maximization. When economists began to look more carefully at the firm, either in agricultural economics or in managerial economics, incentives became central. Indeed, for various reasons, the owner of the firm must delegate various tasks to the members of the firm. This raises first the problems of managing information flows within the firm. This was the first research topic for economists, once they mastered behavior under uncertainty, thanks to Von Neumann and Morgenstern (1943). This line of research culminated in the theory of teams (Marschak and Radner (1972)). This theory recognized the decentralized nature of information, but postulated identical objective functions for the members of the firm considered as a "team". How to coordinate actions among the members of the team by the proper management of information was the central focus of this research. Incentive questions were still outside the scope of the analysis.

However, as soon as one acknowledges that the members of a firm may have different objectives, delegation becomes more problematic as recognized early by Marschak (1955) and also by Arrow (1968) when he observes:

> "by definition the agent has been selected for his specialized knowledge and the principal can never hope to completely check the agent's performance."

Delegation of a task to an agent who has different objectives than the principal who delegates this task is problematic when information about the agent is imperfect. This is the essence of incentive questions. If the agent had a different objective function but no private information, the principal could propose a contract which perfectly controls the agent and induces the latter's actions to be what he would like to do himself in a world without delegation. Again, incentives issues would disappear.

Conflicting objectives and decentralized information are thus the two basic ingredients of incentive theory. That economic agents pursue at least to some extent their private interests is the essential paradigm for the analysis of market behavior by economists. What is proposed by incentive theory is to maintain this major assumption in the analysis of organizations, small numbers markets and any kind of collective decision. This paradigm has its own limits. Social behavior, in particular in small groups, is more complex, and norms of behavior culturally inculcated play a large role in shaping societies. However, it would be foolish not to recognize the role of private incentives in motivating behavior in addition to these cultural phenomena. The purpose of this book is to synthesize what we have learned from the incentives paradigm.[1]

We hope that the step by step approach taken here, as well as our attempt to present many different results in a unified framework, will help the readers not only to know about incentive theory, but to appropriate this indispensable tool for thinking about society.

The starting point of incentive theory corresponds therefore to the problem of delegation of a task to an agent with private information. This private information can be of two types : either the agent can take an action unobserved by the principal, the case of moral hazard or hidden action ; or the agent has some private information about its cost or valuation that is ignored by the principal, the case of adverse selection or hidden knowledge. The theory studies when this private information is a problem for the principal, and what is the optimal way for the principal to cope with it. Another type of information problem has also been raised in the literature, the case of nonverifiability where the principal and the agent share ex post the same information but no third party and, in particular, no Court of Justice can observe this information. One can study to which extent this nonverifiability of some piece of information is problematic for contractual design.

[1] How do private incentives interact with cultural norms of behavior might be the next important step of research needed to be able to offer sensible advice on the design of institutions. It is our conviction nevertheless that for such a goal the mastering of incentive theory is a must.

We will discover that, in general, these informational problems prevent society from achieving the first best allocation of resources which could be possible in a world where all information is common knowledge. The additional costs that must be incurred because of the strategic behavior of privately informed economic agents can be viewed as one category of the transaction costs emphasized by Williamson (1975). They do not exhaust all possible transaction costs, but economists have been rather successful during the last thirty years, in modeling and analyzing this type of transaction costs, providing a good understanding of the limits put by these new costs for the allocation of resources. This work shows that the design of proper institutions for successful economic activities is more complex than one could have thought. This whole line of research provides also a whole set of insights on how to proceed to take into account agents' responses to the incentives provided by institutions.

As the next chapter will illustrate, incentive theory was pervasive in many areas of economics, even though it was not central in economic thinking. Before describing how we will proceed to present this theory, it may be worth mentioning how the major achievement of economics, namely the general equilibrium theory, met incentives.

General equilibrium theory proved apt to powerful generalizations and able to deal with uncertainty, time, externalities, extending the validity of the *invisible-hand* as long as the appropriate competitive markets could be set up. However, at the beginning of the seventies, works by Akerlof (1970), Spence (1974), and Rothschild and Stiglitz (1976) showed in various ways that asymmetric information was posing a much greater challenge, and could not satisfactorily be imbedded in a proper generalization of the Arrow-Debreu theory. The problems encountered were so serious that a whole generation of general equilibrium theorists gave up momentarily the grandiose framework of GE to reconsider the problem of exchange under asymmetric information in its simplest form, i.e., between two traders, and in a sense go back to basics. They joined another group trained in game theory and in the theory of organizations to build the theory of incentives, that we take as encompassing contract theory, principal-agent theory, agency theory and mechanism design.

We will present incentive theory in three progressive steps. Volume I is the first step, in which we consider the principal-agent model where the principal delegates an action to a single agent through a take-it-or-leave-it offer of a contract.

Two implicit assumptions are made here. First, by postulating that it is the principal who makes a-take-it-or-leave-it contract offer to the agent, we put aside the bargaining issues which is a topic for game theory.[2] Second, we assume also the availability of a benevolent Court of Justice which is able to enforce the contract and to impose penalties

[2]See for example Osborne and Rubinstein (1993).

if one of the contractual partners adopts a behavior which deviates from the that one specified in the contract.[3]

Three types of information problems will be considered, adverse selection, moral hazard and nonverifiability. Each of those informational problems leads to a different paradigm and, possibly, to a different kind of agency costs. On top of the usual technological constraints of neoclassical economics, these agency costs incorporate the informational constraints faced by the principal at the time of designing the contract.

In this volume, we will assume that there are no restrictions on the contracts that the principal can offer. As a consequence, the design of the principal's optimal contract reduces to a simple optimization problem.[4] This simple focus will turn out to be already enough to highlight the various trade-offs between allocative efficiency and the distribution of information rents arising under incomplete information. The mere existence of informational constraints may generally prevent the principal from achieving allocative efficiency. The main thrust of the analysis undertaken in this volume is therefore the characterization of the allocative distortions that the principal finds desirable to implement in order to mitigate the impact of informational constraints.

Volume II will be the second step of our analysis. We will consider there situations with one principal and several agents, still without any restriction on the principal's contracts. Asymmetric information may not only affect the relationship between the principal and each of his agents, but it may also plague the relationships between agents. Moreover, pursuing the hypothesis that agents adopt an individualistic behavior, those organizational contexts require a new equilibrium concept, the Bayesian Nash equilibrium, which describes the strategic interaction between agents under incomplete information. Three main themes arise in this context. First, the organization may have been built to facilitate a joint decision between the agents. In such a context, the principal must overcome the free-rider problems which might exist among agents when they must undertake a collective decision. Second, the principal may attempt to benefit from the competition between the agents to relax the informational constraints and better reduce the agents' information rents. Auctions, tournaments, yardstick competition and supervision of an agent by another one are all mechanisms designed by the principal with this purpose in mind. Third, the mere attempt by the principal to use competition between agents may also trigger their collusion against the principal. The principal must now worry not only

[3]Let us stress here the importance of this assumption which is apparently innocuous because, in equilibrium, no penalty is ever paid and the role of the court looks minimal in what follows. However, judges may have to be given proper incentives to enforce contracts. We rely here on the idea that in repeated relationships the desire to maintain their reputation will provide the appropriate incentives. This implicit assumption is a little bit problematic since once could also appeal to the same reputation argument to justify that the principal-agent relationship may achieve allocative efficiency. It will be relaxed in Volume III.

[4]Hence, solving for the optimal contract requires only the simple tools of optimization theory.

about individual incentives, but also about group incentives in a multi-agent organization.

Volume III will be the third step and will analyze the implications of various imperfections in the design of contracts: Informed principal, limited commitment, renegotiation, imperfect coordination among various principals, incomplete contracting on the value of trade. The dynamics of some of these imperfect contractual relationships call for the extensive use of another equilibrium concept: the Bayesian perfect equilibrium. Equipped with this tool, we will be better able to describe the allocation of resources resulting from such imperfect contractual relationships.

In Volume I, we proceed as follows. Chapter 1 gives a brief account of the history of thought concerning incentive theory. It will show that incentives questions have been present in many areas of economics over the last century even though it is only recently that their importance has been recognized and that economists have undertaken a systematic treatment of these issues. Chapter 2 presents the basic rent extraction-efficiency trade-off which arises in principal-agent models with adverse selection. Extensions of this framework to more complex environments are discussed in Chapter 3. Chapter 4 presents the two types of trade-offs under moral hazard: the trade-off between the liability rent extraction and allocative efficiency and the trade-off between insurance and efficiency. Again, extensions of this basic framework are discussed in Chapter 5. Chapter 6 considers the nonverifiability paradigm which in general does not call for economic distortions. Mixed models with adverse selection, moral hazard and nonverifiability are the subject of Chapter 7. The extension of principal-agent models with adverse selection and moral hazard to dynamic contexts with full commitment is given in Chapter 8. Finally, Chapter 9 discusses a number of simple extensions of the basic framework used all over the book.

Chapter 1

Incentives in Economic Thought

1.1 Introduction

Incentive theory emerges with the division of labor and exchange.[1] The division of labor induces the need for delegation and the first historical contracts appear probably in agriculture when a landlord contracts with his tenant. It is then no wonder that Adam Smith encountered incentive problems in his discussion of sharecropping contracts (Section 1.2). Delegation was also needed within firms, hence the importance of the topic in the theory of organizations (Section 1.3).

For private goods, competitive markets ensure efficiency despite the decentralized nature of the information about individuals' tastes and firms' technologies. Implicitly, yardstick competition solves adverse selection problems and the fixed-price contracts associated with exogenous prices solve moral hazard problems. However, markets fail for pure public goods and public intervention is thus needed. In this case, the mechanisms used for those collective decisions must solve the incentive problem of acquiring the private information that agents have about their preferences for public goods (Section 1.4). Voting mechanisms are particular incentive mechanisms without any monetary transfers for which the same question of strategic voting, i.e, not voting according to the true preferences, can be raised (Section 1.5).

For private goods, increasing returns to scale create a situation of natural monopoly far away from the world of competitive markets. When the monopoly has private information about its cost or demand, its regulation by a regulatory commission becomes a principal-agent problem (Section 1.6).

Exchange raises incentive issues when the commodity which is bought has a value

[1] Actually, one could also argue that incentive issues arise within the family if one postulates different objective functions for the members of the family.

unknown to the buyer but known to the seller. It is the case, in particular, in insurance markets when the insurance company buys a risk plagued with moral hazard or adverse selection. The insurance company faces a principal-agent problem with each insured agent, but may nevertheless have a statistical knowledge of the distribution of risks (Section 1.7). A similar situation occurs when a government attempts to redistribute income between wage earners of different and unknown productive abilities (Section 1.8) or when a monopolist looks for the optimal discriminating contract to offer to a population of consumers with heterogenous tastes for its product (Section 1.9). Of course, incentive issues were encountered in managing socialist economies as profit incentives of managers were suppressed by public ownership of the means of production (Section 1.10). The idea that, in non-competitive economies, it is necessary to design mechanisms taking into account communication and incentives constraints was further developed by theorists dealing with non convex economies and this led to the mechanism design methodology (Section 1.11). The mechanism design methodology provides a useful tool to understand the allocation of resources in multi-agent frameworks when information is decentralized. A natural field to apply this methodology is the theory of auctions. Auctions are indeed mechanisms used by principals to benefit from the competition among several agents (Section 1.12).

1.2 Adam Smith and Incentive Contracts in Agriculture

In his discussion of the determination of wages (Chapter VII, Book I in Smith (1776)), Adam Smith recognized the contractual nature of the relationship between the masters and the workmen. He put forward the conflicting interests of those two players and already recognized that the bargaining power was not evenly distributed between them, the master having in general all the bargaining power. In the modern language of the Theory of Incentives, the masters are principals and the workmen their agents.

> "What are the common wages of labour, depends everywhere upon the contract usually made between those two parties,whose interests are not the same. The workmen desire to get as much, the masters to give as little as possible." p. 66

Smith also stressed one of the basic constraints that we model later on: The agent's participation constraint which limits what the principal can ask from the agent:

> "A man must always live by his work, and his wages must at least be sufficient to maintain him." p. 67

Smith did not have a vision of economic actors as long-run maximizers of utility. He worried about the consequences of high-power incentives for short-run maximizers.

> "Workmen, [...], when they are liberally paid by the piece, are very apt to overwork themselves, and to ruin their health and constitution in a few years." p. 81

He stressed the lack of appropriate incentives for slaves:

> "the work done by slaves, though it appears to cost only their maintenance, is in the end the dearest of any. A person who can acquire no property, can have no other interest but to eat as much, and to labour as little as possible." p. 365

To explain the survivance of such highly inefficient contracts, Adam Smith also appealed to non-economic motives:

> "The pride of man makes him love to domineer, and nothing mortifies him so much as to be obliged to condescend to persuade his inferiors." p. 365

Smith's most precise and famous discussion of incentives appears in Chapter II, Book III, when he wants to explain the discouragement of agriculture in the ancient state of Europe after the fall of the Roman Empire. He describes the status of metayers (Coloni Partarii in Ancient Time, steel-bow tenants in Scotland):

> "The proprietor furnished them with the seed, cattle and instruments of husbandry. The produce was divided equally between the proprietor and the farmer." p. 366

However, Smith did not conclude that metayers will not exert the appropriate level of effort to maximize social value, as modern incentive theory would claim.

> "Such tenants, being free men, are capable of acquiring property, and having a certain proportion of the produce of the land, they have a plain interest that the whole produce would be as great as possible, in order that their own proportion may be so." p. 366

At several place in this volume, we will see the fundamental trade-off between incentive and the distribution of the gains from trade. Clearly Smith was not aware of this trade-off.

Rather, he saw the most serious incentive problems in the absence of invesment in the land by tenants and the unobservable misuse of instruments of husbandry provided by the proprietor.

> "It could never, however, be the interest even of this last species of cultivators (the metayers) to lay out, in the further improvement of the land, any part of the little stock they might save from their own share of the produce, because the lord, who laid out nothing, was to get one-half of whatever it produced... It *might* be the interest of metayer to make the land produce as much as could be brought out of it by means of the stock furnished by the proprietor; but it could never be in his interest to mix any part of his own with it. In France,..., the proprietors complain that their metayers take every opportunity of employing the master's cattle rather in carriage than in cultivation; because in the one case they get the whole profits for themselves, in the other they share them with their landlords." p. 367

Note the ambiguous *"might"*, which shows that Smith envisioned probably under-effort but that he considered it as secondary compared to the under-investment effect. However, the alternative use of cattle is a typical example of what we will call a hidden action problem or a moral hazard problem.

Smith's criticism of sharecropping has been the point of departure of a large literature in agricultural economics, in history of thought and in economic theory trying to understand the characteristics of sharecropping contracts. Following A. Smith and until Johnson (1950), economists have considered sharecropping to be a "practice which is hurtful to the whole society", an unexplained failure of the indivisible hand that should be either discouraged by taxation or improved by appropriate sharing of variable factors.[2] A better understanding of the phenomenon was only achieved when the economists reconsidered the problem equipped with the principal-agent theory.[3]

1.3 Chester Barnard and Incentives in Management

As we saw above Smith (1776) already discussed the problems associated with piece-rate contracts in the industry. Babbage (1835) made a further step by understanding the need

[2]See Schickele (1941) and Heady (1947).
[3]See Stiglitz (1974).

for precise measurement of performances to set up efficient piece-rate or profit-sharing contracts.

> "It would, indeed, be of great mutual advantage to the industrious workman, and to the mastermanufacturer in every trade, if the machines employed in it could register the quantity of work which they perform, in the same manner as a steam-engine does the number of strokes it makes. The introduction of such contrivances gives a greater stimulus to honest industry than can readily be imagined, and removes one of the sources of disagreement between parties." p. 297

Also, Babbage proposed various principles to remunerate labor:

> "The general principles on which the proposed system is founded, are
>
> 1. That a considerable part of the wages received by each person should depend on the profits made by the establishment; and,
> 2. That every person connected with it should derive more advantage from applying any improvement he might discover than he could by any other course."
>
> Babbage (1989, Vol. 8, p. 177).

However, Barnard (1938) can probably be credited of the first attempt to define a general theory of incentives in management, with Chapter 11 —the economy of incentives— and Chapter 12 —the theory of authority— of his celebrated book "The Function of the Executive" that he wrote after a long career in management, in particular as President of the New Jersey Bell Telephone Company:

> "an essential element of organizations is the willingness of persons to contribute their individual efforts to the cooperative system... Inadequate incentives mean dissolution, or changes of organization purpose, or failure to cooperate. Hence, in all sorts of organizations the affording of adequate incentives becomes the most definitely emphasized task in their existence. It is probably in this aspect of executive work that failure is most pronounced." p. 139

Actually, Barnard had a large view of incentives, involving both what we would call nowadays monetary and non-monetary incentives:

> "An organization can secure the efforts necessary to its existence, then, either by the objective inducements it provides or by changing states of mind ... We shall call the process of offering objective incentives "the method of incentives"; and the processes of changing subjective attitudes "the method of persuasion"." p. 142

The incentives may be specific or general.

> "The specific inducements that may be offered are of several classes, for example: a) material inducements; b) personal non material opportunities; c) desirable physical conditions; d) ideal benefactions. General incentives afforded are, for example: e) associational attractiveness; f) adaptation of conditions to habitual methods and attitudes; g) opportunity of enlarged participation; h) the condition of communion." p. 142

Barnard also stressed the ineffectivity of material incentives so far almost exclusively considered by economic theory:

> "even in purely commercial organization material incentives are so weak as to be almost negligible except when reinforced by other incentives." p. 144
>
> "Persuasion includes: a) the creation of coercive conditions (as forced exclusion of indesirables); b) the rationalization of opportunities (if the conviction that material things are worth while... succeeds in capturing waste effort and wasted time... it is clearly advantageous); c) the unculcation of motives." p. 154

Barnard pointed out the necessary delicate balance of the various types of incentives for success. Furthermore, such a good balance is highly dependent of an unstable environment (through competition in particular) and of the internal evolution of the organization itself (growth, change of personel). Finally, in his chapter on authority, Barnard recognized that incentive contracts do not rule all the activities within an organization. The distribution of authority along communication channels is also necessary to achieve coordination and promote cooperation.

> "Authority arises from the technological and social limitations of cooperative systems on the one hand, and of individuals on the other." p. 184

In modern language, he is saying that the incompleteness of contracts and the bounded rationality of members in the organization require that some leaders be given authority to decide in circumstances not anticipated precisely by the contracts. His main point is then to stress the need to satisfy ex post participation constraints of members who accept non contractual orders only if they are compatible with their own long-run interests.

> "A person can and will accept a communication as authoritative only when..., at *the time of his decision*, he believes it to be compatible with his personal interest as a whole." p. 165

Barnard's work emphasized the need to induce appropriate effort levels from members of the organization -the moral hazard problem- and to create authority relationships within the organization to deal with the necessary incompleteness of incentive contracts. We will then have to wait for Arrow (1963) to introduce in the literature on the control of management the idea of moral hazard borrowed from the world of insurance. This work will be further extended by Wilson (1968) and Ross (1973) who will redefine it explicitly as an *agency problem*. The chapter on authority written by Barnard directly inspired Simon (1951)'s formal theory of the employment relationship. Finally, Williamson (1975) followed Barnard and Simon to develop his transaction costs theory for the case of symmetric but nonverifiable information between two parties.[4]

Grossman and Hart (1986) modeled this paradigm and this led to the large recent literature on incomplete contracts.[5]

1.4 Hume, Wicksell, Groves: The Free Rider Problem

Hume (1740) may be credited of the first explicit statement of the *"free-rider problem"*.

> "Two neighbours may agree to drain a meadow, which they possess in common; because it is easy for them to know each others mind; and each must perceive, that the immediate consequence of his failing in his part, is the abandoning the whole project. But it is very difficult, and indeed impossible, that a thousand persons shou'd agree in any such action; it being difficult for them to concert so complicated a design, and still more difficult for them to execute it; while each seeks a pretext to free himself of the trouble and expence, and wou'd lay the whole burden on others." p. 538

[4] See Williamson's citation in Section 6.1.

[5] See Hart (1995) for a recent synthesis.

At the end of the 19th century, a lively debate over public finance took place among European economists between the "benefit" approach and the "ability to pay" approach to taxation. In particular, Mazzola, Pantaleoni, de Viti de Marco in Italy, Sax in Austria used the "modern" concepts of marginal utility and subjective value, extending the benefit approach implicit in the writings of many authors of the 18th century, such as Bentham, Locke and Rousseau. Wicksell (1896), in his discussion of Mazzola's contribution, pointed out what became known later as the free-rider problem, which had been ignored in the benefit approach to taxation.

> "If the individual is to spend his money for private and public uses so that his satisfaction is maximized he will obviously pay nothing whatsovever for public purposes... Whether he pays much or little will affect the scope of public service so slightly, that for all practical purposes, he himself will not notice it at all. Of course, if everyone were to do the same, the State will soon cease to function." p. 81

Wicksell suggested a solution: The principle of (approximative) unanimity and voluntary consent. Each item in the public budget must be voted simultaneously with the determination of its financing and must be accepted only if unanimity (or quasi-unanimity) is obtained.[6] If we could ignore strategic behavior, this process would lead to Pareto optimality. However, which one of the Pareto optima will be reached depends upon the sequential realization of the decision-making process. Indeed, this is the main reason justifying strategic behavior by the participants as they try to manipulate the path of the procedure.

With the exception of Bowen (1943)'s voting procedure discussed in the next section, nothing was proposed until the seventies to solve the free-rider problem which appeared really formidable. Nevertheless, in 1971, Drèze and de la Vallée Poussin extended to public goods the literature on iterative planning procedures of the sixties. At each step of the procedure, agents announce their marginal rates of substitution between public goods and private good. They noted that revelation of the true marginal rates of substitution is a maximin strategy, a weak incentive property.

Finally, Clarke (1971), Groves (1973), Groves and Loeb (1975), making strong restrictions on preferences to evade the Gibbard-Satterthwaite *"Impossibility Theorem"*,[7] provided mechanisms with monetary transfers inducing truthful revelation of preferences and making the Pareto optimal public good decision. The literature which followed[8] developed substantially incentive theory and the mechanism design methodology.

[6]This notion was later formalized by Foley (1967).
[7]See Section 1.5 below.
[8]See Green and Laffont (1979) and Aspremont and Gérard-Varet (1979).

1.5 Borda, Bowen, Vickrey: Incentives in Voting

Since the beginning of the theory on voting, the issue of strategic voting was noticed. Borda (1781) recognized it when he proposed his famous *Borda rule*

"My scheme is only intented for honest men."

We have to wait for Bowen (1943) to see a first attempt at addressing the issue of *"strategic voting"*. For allocating public goods, Bowen (as we mentioned in Section 1.4) was searching in voting an alternative to the missing expression of preferences in markets that exists for private goods. He realized the difficulty of strategic voting:

> "At first sight it might be supposed that this information could be obtained from his vote... But the individual could not vote intelligently, unless he knew in advance the cost to him of various amounts of the social good, and in any case the results of voting would be unreliable if the individual suspected that his expression of preference would influence the amount of cost to be assessed against him." Bowen (1943, p. 129 in Arrow and Scitovsky (1969)).

Bowen assumed that the distribution of the cost of the public good was exogenously fixed (for example equal sharing of cost) and considered successive votes on increments of the public good. He observed that at each step it is in the interest of each voter to vote yes or no according to his true preferences. Such a procedure converges to the optimal level of public good if agents are myopic and consider only their incentives at each step.[9] Single-peaked Black (1948), years after Borda, Condorcet, Laplace and Dogson, reconsidered the theory of voting and exhibited a wide class of cases (single-peaked preferences) for which majority voting leads to transitivity of social choice, a solution to the 1785 Condorcet paradox. He eliminated, by assumption, strategic issues.

> "When a member values the motions before a committee in a definite order, it is reasonable to assume that, when these motions are put against each other, he votes in accordance with his valuation." Black (1948, p. 134 in Arrow and Scitovsky (1969)).

When Arrow (1951) founded the formal theory of social choice by proving that there is no "reasonable" voting method yielding to a non dictatorial social ranking of social alternatives which avoids intransitivity when no restriction is placed on individual preferences, he also abstracted from the gaming issues and noticed:

[9] See Green and Laffont (1979, Chapter 14) for a more detailed analysis of this procedure.

> "The point here, broadly speaking, is that, once a machinery for making social choices from individual tastes is established, individuals will find it profitable, from a rational point of view, to misrepresent their tastes by their actions or, more usually, because some other individual will be made so much better off by the first individual's misrepresentation that he could compensate the first individual in such a way that both are better off than if everyone really acted in direct accordance with his tastes."[10] p. 7

In a paper which provides a very lucid exposition of Arrow's impossibility theorem, Vickrey (1960) raised the question of strategic misrepresentation of preferences in a social welfare function which associates a social ranking to individual preferences.

> "There is another objection to such welfare functions, however, which is that they are vulnerable to strategy. By this is meant that individuals may be able to gain by reporting a preference differing from that which they actually hold." p. 517,

and:

> "Such a strategy could, of course, lead to a counterstrategy, and the process of arriving at a social decision could readily turn into a "game" in the technical sense." p. 518

Dummet and Farquharson (1961) will indeed pursue the analysis of such voting games in terms of non-cooperative Nash equilibria. Vickrey (1960) further explained that the social welfare functions which satisfy the assumptions of Arrow's theorem, in particular the independence assumption, are immune to strategy. Then, comes his conjecture acknowledged by Gibbard (1973):

> "It can be plausibly conjectured that the converse is also true, that is, that if a function is to be immune to strategy and to be defined over a comprehensive range of admissible rankings, it must satisfy the independence criterion, though it is not quite so easy to provide a formal proof of this." p. 588.

Therefore, Vickrey is led, through Arrow's theorem, to an impossibility result, namely the strategic manipulability of any method of aggregating individual preferences or of

[10] Note that the last part of this quote refers to incentives for groups.

any voting mechanism. The route toward the impossibility of non-manipulable and non-dictatorial mechanisms via Arrow's theorem was suggested. A complete proof, the greatest achievement of social choice theory since Arrow's theorem, came thirteen years later in Gibbard (1973).[11] The importance of Gibbard's theorem for incentive theory lies in showing that with no prior knowledge of preferences, non-dictatorial collective decision methods cannot be found where truthful behavior is a dominant strategy. The positive results of incentive methods in practice will have to be looked for in restrictions on preferences, as in the principal-agent theory, or in the relaxation of the required strength of incentives by giving up dominant strategy implementation.

1.6 Léon Walras and the Regulation of Natural Monopolies

Walras (1897) defined a natural monopoly as an industry where monopoly is the efficient market structure and suggested, following A. Smith (1776), to price the product of the firm by balancing its budget. This led to the Ramsey (1927) and Boiteux (1956) theory of optimal pricing under a budget constraint.

After some price cap regulation attempts in the 19th century, the practice of regulation was rate of return regulation which ensures prices covering costs inclusive of a (higher than the market) cost of capital. This led to the Averch and Johnson (1962) over-capitalization result largely overemphasized.

In 1979, Loeb and Magat finally put the regulation literature in the framework of the principal-agent literature with adverse selection by stressing the lack of information of the regulator. They proposed to use a Groves dominant strategy mechanism which solves the problem of asymmetric information at no cost when there is no social cost in transfers from the regulator to the firm.

Baron and Myerson (1982) transformed the problem into a second-best problem by weighting the firm's profit with a smaller weight than consumers' surplus in the social welfare function maximized by the regulator. Then, optimal regulation entails a distortion from the first-best (pricing higher than marginal cost) to decrease the information rent of the regulated firm. Laffont and Tirole (1986) used a utilitarian social welfare function with the same weight for profit and consumers' surplus, but introduced a social cost for public funds (due to distortive taxation) which creates also a rent-efficiency trade-off. Their model features both adverse selection and moral hazard, but the ex post observability of cost (commonly used in regulation) makes it technically an adverse selection model.[12]

---

[11] See also Satterthwaite (1975).

[12] See Chapter 7 below.

This model is developed in Laffont and Tirole (1993) along many dimensions (dynamics, renegotiation, auctions, political economy...).

1.7 Knight, Arrow, Pauly: Incentives in Insurance

The notion of moral hazard, i.e., the ability of insured agents to affect the probabilities of insured events was well known in the insurance profession.[13] However, the insurance writers tended to look upon this phenomenon as a moral or ethical problem affecting their business.

In 1963, Arrow introduced this concept in the economics literature [14] and argued that it led to a market failure as some insurance markets would not emerge due to moral hazard. Arrow was quite influenced by the moral connotation of the concept and looked for solutions involving changes of ethical attitudes. Pauly (1968) rejected this approach, by arguing that it was quite natural for agents to react to zero price —like demanding more health consumption if health was free— and that the non-insurability of some risks did not imply a market failure as no proof of the superiority of public intervention faced with the same informational problems was given. Pauly (1974) and Helpman and Laffont (1975) showed that indeed competitive insurance markets (with linear prices) were inefficient in the sense that an uninformed government could improve upon the free market outcome.

Spence and Zeckhauser (1971) looked for more sophisticated contracts (non-linear prices) by solving the maximization of the welfare of a representative agent with a break-even constraint for the insurance company and the moral hazard constraint that each agent chooses its level of self-protection optimally. When the self-protection variable is chosen before nature selects the states of nature (i.e., who has an accident, who does not), they obtained the moral hazard variable model with a continuum of agents and a break-even constraint. When the self-protection variable is chosen after nature selects the

[13]See for example Faulkner (1960) and Dickerson (1957).

[14]Leroy and Singell (1987) make the claim we share that, by uncertainty, Knight (1921) meant situations in which insurance markets collapse because of moral hazard or adverse selection.

> "The classification or grouping (necessary for insurance) can only to a limited extent be carried out by any agency outside the person himself who makes the decisions, because of the peculiarly obstinate connection of a moral hazard with this sort of risks." p. 251
> "We have assumed ... that each man in society knows his own powers as entrepreneur, but that men know nothing about each other in this capacity... The presence of true profit, therefore, depends... on the absence of the requisite organization for combining a sufficient number of instances to secure certainty through consolidation. With men in complete ignorance of the powers of judgement of other men it is hard to see how such organization can be effected." p. 284

However, Knight did not recognize that problems of moral hazard and adverse selection can be attenuated or eliminated with properly structured contracts.

states of nature, they have both moral hazard and adverse selection, making the problem quite close to the Mirrlees optimal income tax problem [15] (as already noted by Zeckhauser (1970)).

Ross (1973) expressed the pure principal-one agent model with only moral hazard and an individual rationality constraint for the agent, before it received its modern treatment in Mirrlees (1975), Guesnerie and Laffont (1979), Holmström (1979), Shavell (1979) and later in Grossman and Hart (1983).

The Pareto inefficiency of competitive insurance markets (with linear prices) with adverse selection was shown in Rothschild and Stiglitz (1976)[16] and their successors studied various forms of competition in non-linear tariffs. As in the case of moral hazard, one can also study the optimal non-linear tariff which maximizes the expected welfare of a population of agents having private information about their own risk characteristics.[17] However, this problem was encountered earlier in the literature on price discrimination with quality replacing quantity.[18]

1.8 Sidgwick, Vickrey, Mirrlees: Redistribution and Incentives

The separation of efficiency and redistribution in the second theorem of welfare economics rests on the assumption that lump-sum transfers are feasible. As soon as the bases for taxation can be affected by agents' behavior, deadweight losses are created. Then, raising money for redistributive purposes destroys efficiency. More redistribution requires more inefficiency. A trade-off appears between redistribution and efficiency. When labor income is taxed, the leisure-consumption choices are distorted and the incentives for work are decreased. There exists a *redistribution-incentives trade-off*. Sidgwick (1883) in his Method of Ethics was apparently the first writer to recognize the incentive problems of redistribution policies.

> "It is conceivable that a greater equality in the distribution of products would lead ultimately to a reduction in the total amount to be distributed in consequence of a general preference of leisure to the results of labor." Chapter 7, Section 2.

[15] They do not go much beyond writing first-order conditions for this problem, and refer to Mirrlees (1971) when they use the Pontryagin principle.

[16] See also Akerlof (1970) and Spence (1973).

[17] See Stiglitz (1977).

[18] See Mussa and Rosen (1978) and Guesnerie and Laffont (1984) for modern treatments.

The informational difficulty associated with income taxation is that the supply of labor is not observable and therefore not controllable, hence the distortion. However, if the wage was observable, as well as income, the supply of labor would be easily recovered. The next stage in the modeling of the problem was to assume that wages which equate innate abilities are private information of the agents.[19] Income, the observable variable, is the product of a moral hazard variable, the supply of labor, and of an adverse selection variable, ability.

A major step was achieved by Vickrey, who had been senior economist of the tax research division of the US Treasury Department and tax expert of the governor of Puerto Rico. As early as 1945, he used the insights of Von Neumann and Morgenstern to model the optimal income tax problem as a principal-agent problem where the principal is the tax authority and the agents the tax payers. In Vickrey (1945) he defined the objective function of the government:

> "If utility is defined as that quantity the mathematical expression of which is maximized by an individual making choices involving risk, then to maximize the aggregate of such utility over the population is equivalent to choosing that distribution of income which such an individual would select were he asked which of various variants of the economy he would become a member of, assuming that once he selects a given economy with a given distribution of income he has an equal chance of landing in the shoes of each member of it." p. 329

Equipped with this utilitarian social welfare criterion, with, in passing, the Harsanyi (1955) interpretation of expected utility as a justice criterion, he formulated the fundamental problem of optimal income taxation:[20]

> "It is generally considered that if individual incomes were made substantially independent of individual effort, production would suffer and there would be less to divide among the population. Accordingly some degree of inequality in needed in order to provide the required incentives and stimuli to efficient cooperation of individuals in the production process." p. 330

> "The question of the ideal distribution of income, and hence of the proper progression of the tax system, becomes a matter of compromise between equality and incentives." p. 330

[19]Note here a difficulty. Wages are paid by employers who must know ability. Implicitly, collusion between employers and workers is assumed. With a profit tax it is easy to fight this type of collusion.

[20]Vickrey viewed his work as a generalization of Edgeworth's minimum-sacrifice principle (1897). Also, Edgeworth's optimal indirect taxes can be viewed also as an incentive problem.

He then proceeded to a formalization of the problem which is still the current one. The utility function of any individual is made a function of his consumption and of his productive effort. There is a relationship between the amount of output on the one hand and the amount of effort and unknown productive characteristics of the individual on the other hand. This leads to an alternative form of the utility function which depends now on consumption, output and the individual's characteristics. Taxation creates a relationship between output and consumption. Adjusting his effort or output optimally, the individual obtains his supply of effort characterized by a first-order condition which is the first-order condition of incentive compatibility for an adverse selection problem. He stated the government's optimization problem which is to maximize the sum of individuals' utilities under the incentive compatibility conditions and the budget equation of the government. Recognizing a calculus of variation problem, he wrote the Euler equation and gave up:

> "Thus even in this simplified form the problem resists any facile solution." p. 332

The Pontryagin principle was still far away and twenty six years will be needed to reach Mirrlees (1971)'s neat formulation and solution of the problem.[21]

Note that the problem analyzed here is not stricto sensu a delegation problem as we defined it above. The principal is actually delegated by the taxpayers the task of redistributing income, i.e., a particular public good problem. The principal observes neither the effort level of a given agent, nor his productive characteristics. However, by observing output which is a function of both, it can reduce the problem to a one dimensional adverse selection problem. The principal is not facing a single agent over the characteristics of which he has an asymmetry of information, but a continuum of them for which he knows only the distribution of characteristics. Nevertheless, the problem is mathematically identical to a delegation problem with a budget balance equation instead of a participation constraint.[22]

1.9 Dupuit, Edgeworth, Pigou: Price Discrimination

When a monopolist or a government wants to extract consumers' surpluses in the pricing of a commodity, it faces in general the problem of the heterogeneity of consumers' tastes. Even if it knows the distribution of tastes, it does not know the type of any given consumer. By offering different menus of price-quality or price-quantity pairs, i.e., by using second-degree price discrimination, the government can increase its objective function. Such an

[21] Zeckhauser (1970) and Wesson (1972) formulated special cases of the optimal incentives-redistribution problem that they solved approximately without being aware of the Vickrey model.

[22] At least when the types of the agents are independently distributed.

anonymous menu is an incentive mechanism which leads consumers to reveal their type by their self-selection in the menu.

Dupuit (1844) developed the concept of consumer surplus and used it to discuss price discrimination. Dupuit was well aware of the incentive problems faced by the pricing of infrastructures.

> "The best of all tariffs would be the one which would make pay those which use a way of communication a price proportional to the utility they derive from using this service... I do not have to say that I do not believe in the possible application of this voluntary tariff; it would meet an insurmountable obstacle in the universal dishonesty of passants, but it is the kind of tariff one must try to approach by a compulsory tariff," Dupuit (1849), p. 223.

Edgeworth (1913) extended the theory for price discrimination for the railways industry. Pigou (1920) characterized the different types of price discrimination. Gabor (1955) discussed block tariffs or two-part tariffs which had been recently introduced in the electricity industry in England and showed that with one type of consumers two part tariffs are equivalent to first degree price discrimination. Oi (1971) derived an optimal two-part tariff. Mussa and Rosen (1978), Spence (1977), Goldman, Leland and Sibley (1984) provided the general framework to derive for a monopolist an optimal tariff which is non-linear in prices or qualities, substantially later than similar work in the income tax or insurance literature.

1.10 Incentives in Planned Economies

We must distinguish between the Soviet practice and the Theory of Planning developed in the western countries. As explained by Berliner (1976, p. 401) "In the early years of the Soviet period there was some hope that socialist society could count on the spirit of public service as a sufficient motivation for economic activity. With the intense industrialization drive of the thirties, however, that hope was gradually abandoned. In a historic declaration in 1931, Stalin renounced the equalitarian wage ethic that had obliterated "any difference between skilled and unskilled work, between heavy and light work"." Following his biting denunciation of "equality mongering", there evolved a new policy in which personal "material incentives" —primarily money incomes— became the major instrument for motivating economic activity.

In the Soviet Union, a general set of managerial incentive structures developed during the thirties and lasted for three decades. In this classical period, the manager's incomes were decomposed in a salary, a basic bonus and the Enterprise Fund. This incentive structure had many defects (problems with new products, no proper incentives for cost minimization, ratchet effect...). It was critized and under constant evolution. With the passing of Stalin, the discussion became more intense and quite open with the 1962 Liberman paper in the Pravda and culminated in the 1965 Reform. A literature studying in detail the new incentive structure developed in the Western world among Soviet specialists.[23] In the famous socialist controversy of the thirties, incentives were largely overlooked. Lange (1936) perceived no problem with imposing rules to managers.

> "The decisions of the managers of production are no longer guided by the aim to maximize profit. Instead, there are certain rules imposed on them by the Central Planning Bureau which aim at satisfying consumers' preferences in the best way possible.
>
> One rule must impose on each production plant the choice of the combination of factors of production and the scale of output which minimizes the average cost of production.
>
> The second rule replaces the free entry of firms into an industry or their exodus from it. This leads to an equality of average cost and the price of the product."

Lerner (1934) pointed out the difficulty arising with a small number of firms having increasing returns to scale and reformulated the rules as: Every producer must produce whatever he is producing at the least total cost, and a producer shall produce any output or any increment of output that can be sold for an amount equal to or greater than the marginal cost of that output or increment of output.[24] Even in 1967, Lange did not see any problem of incentives in the working of the socialist economy. "Were I rewrite my essay today my task would be much simpler. My answer to Hayek and Robbins would be: so what's the trouble? Let us put the simultaneous equations on an electronic computer and we shall obtain the solution in less than a second. The market process with its cumbersome tâtonnements appears old fashioned."[25]

It is therefore not surprising that the voluminous mathematical theory of iterative

[23] Leeman (1970), Keren (1972), Weitzman (1976),...

[24] Note that Lerner is here simply rediscovering Laundhart (1885)'s marginal cost pricing principle that the last author associated with government ownership. This principle will be most clearly articulated by Hotelling (1939).

[25] When, at the end of his life around 1964, Lange recognizes more fully the role of incentives, it is about the innovation process and not the every day life of the planning system.

> "What is called optimal allocation is a second-rate matter, what is really of prime importance is that of incentives for the growth of productive forces (accumulation and progress in technology)".

See Kowalik (1976).

planning developed in the sixties did not pay any attention to incentives.[26] Such a concern appeared only marginally in Drèze and Vallée Poussin (1971), where truthful reporting of private characteristics was shown to be a maximum strategy in a planning procedure for public goods. In 1974, Weitzman, who had participated to the development of the iterative planning literature, made a direct criticism of the implicit idea that the planning with prices was good for incentives.

> "It seems to me that a careful examination of the mechanisms of successive approximation planning shows that there is no principal informational difference between iteratively finding an optimum by having the center name prices while the firm responds with quantities, or by having the center assign quantities while the firm reveals costs or marginal costs."

Considering then an explicit planning problem with asymmetric information, he compares price mechanisms and quantity mechanisms. This will be the point of departure of the more general approach in terms of nonlinear prices by Spence (1976). From then on, planning procedures were more systematically studied from the point of incentives.[27] However, by then, the lack of interest for iterative planning was fairly general.

1.11 Leonid Hurwicz and Mechanism Design

When general equilibrium theorists attempted to extend the resource allocation mechanisms to non convex environments they realized that new issues of communication and incentives arose.

> "In a broader perspective, these findings suggest the possibility of a more systematic study of resource allocation mechanisms. In such a study, unlike in the more traditional approach, the mechanism becomes the unknown of the problem rather than a datum...
> The members of such a domain (of mechanisms) can then be appraised in terms of their various "performance characteristics" and, in particular, of their (static and dynamic) optimality properties, their informational efficiency, and the compatibility of their postulated behavior with self-interest (or other motivational variables)." Hurwicz (1960, p. 62) in Arrow and Scitovsky (1969)

[26] See Heal (1973) for a synthesis.
[27] See Laffont (1985) for a survey.

Hurwicz (1960) dedicated his paper to Jacob Marschak. Indeed Marschak was the only major economist aware of incentive problems in the fifties, problems that he chose not to study.

> "This raises the problem of incentives. Organization rules can be devised in such a way that, if every member pursues his own goal, the goal of the organization is served. This is exemplified in practice by bonuses to executives and the promises to loot to besieging soldiers; and in theory, by the (idealized) model of the laisser faire economy. And there exist, of course, also negative incentive (punishments).
> I shall have to leave the problem of incentives aside," Marschak (1955).

Marschak was familiar with the literature of statisticians who became aware of incentive problems quite early. The problem of moral hazard arose in sampling theory for quality control. Whittle (1954) and Hill (1960) understood that the distributions of quality were endogenous and dependent on the care taken in the production process. They studied how to take into account this non controllable effort level in their analysis of quality from a sample. Adverse selection appeared when forecasting probabilities of some events. Good (1952), McCarthy (1956) and later Savage (1971) looked for payment formulas leading forecasters to announce their true estimated probabilities and discovered the incentive constraints for the revelation of information.

Economists around Hurwicz developed a general framework, the mechanism design approach, which treated the competitive markets as just one particular institution in a much more general family of mechanisms run by benevolent planners. During the sixties the emphasis of the research was on the communication costs required by non conventional environments until Groves (1973), influenced by Schultze (1969);[28] called for considering incentives in public policy and constructed incentive compatible mechanisms in a team problem.

The next major step was the understanding of the Revelation Principle[29] which shows that, with adverse selection and moral hazard, any mechanism of organizing society is equivalent to an incentive compatible mechanism by which all informed agents reveal their private information to a planner who recommends actions.[30] The Revelation Principle provides the appropriate framework for the normative analysis of economies with

[28] Schultze wrote, p. 151. "public action need not be simply the provision of public facilities... to offset the economic losses caused by private actions. Rather the objectives of public policy, in such cases, should include a modification of the "signals" given and incentives provided by the market place so as to induce private actions consistent with public policy."

[29] See Gibbard (1973), Green and Laffont (1977), Dasgupta, Hammond and Maskin (1978) and Myerson (1979).

[30] Maskin (1977)'s Nash implementation theorem is the major result when a principal designs a mechanism to be played by agents who know their respective characteristics.

asymmetric information and contracts which can be written on all observable variables. It delivers a neat methodology to study incentive theory that we will use in most of the book.

1.12 Auctions

Auctions are mechanisms by which principals attempt to use the competition among agents to decrease the information rents they have to give up to the agent they are contracting with. It requires a modeling of the relationship between bidders (the agents) who bid under incomplete information about the other agents' valuations for the auctioned good or contract.

Even though auctions have been used at least as far back as 500 BC in Babylon, the first academic work on auctions seems to date from 1944 with a thesis on competitive bidding for securities in which Friedman (1956) presented a method to determine optimal bids in a first-price, sealed-bib auction. In this operation research approach he assumed that there was a single strategic bidder. Vickrey (1961) in a monumental paper provided the first equilibrium theoretic analysis of the first price auction that he compared to the second price auction, often called the Vickrey auction.

It is only after the clarification of the Bayesian Nash equilibrium concept by Harsanyi (1967, 1968) that the theory of auctions was massively developed. Three major models were particularly developed. The independent value model due to Vickrey (1961), the symmetric common value model due to Rothkopf (1969) and Wilson (1969, 1977) and the symmetric common value model due to Wilson (1967, 1969). In a major synthetic paper Milgrom and Weber (1982) showed that most of these models are special cases of the affiliated value paradigm and they clarified the winner's curse developed at the occasion of empirical work about auctions for oil drilling rights in the Gulf of Mexico (Capen et alii (1971)). Myerson (1981) used the general mechanism approach to characterize the optimal auctions in models with private values.