**Jean-Jacques Laffont
and David Martimort**

The Theory of Incentives

The Principal-Agent Model

激励理论
委托代理模型

The Theory of Incentives

THE PRINCIPAL-AGENT MODEL

Jean-Jacques Laffont and

David Martimort

As the economy of incentives as a whole in terms of organization is not usually stressed in economic theory and is certainly not well understood, I shall attempt to indicate the outlines of the theory.

Chester Barnard (1938)

Foreword

The development of the theory of incentives has been a major advance in economics in the last thirty years. The objective of this book is to provide easy access to this theory for undergraduate and first-year graduate students in economics. Our goal is not to be as complete as possible in covering and surveying the many contributions that have flourished in the realm of incentive theory. Instead, our contribution is methodological and intended to offer students some initial clues for analyzing the issues raised by this theory. As much as possible we have favored the simplest models to explain the core of the theory. The exposition has been divided into three books for methodological clarity. This volume presents the basic principal-agent theory with complete contracts. It allows a first exposition of the transaction costs created by contracting under asymmetric information without having to appeal to sophisticated game theory concepts.

The book allows for two levels of reading. Certain sections in some chapters are marked with a star to the right of the section head. A first reading should concentrate on the non-starred sections. These sections are accessible to readers who have a simple knowledge of maximization with inequality constraints. Most essential economic lessons can be understood from this first level of reading. Covering the starred sections will enable students to manipulate the concepts for better

assimilation and will put them in contact with the advanced topics of the field. Going through the text and checking the proofs should provide a good way for students to appropriate the material. Throughout the book we have listed under the heading 📖 the major references that are useful to pursue the study of incentive theory.

We thank our students whose excitement for the topic has led to this enterprise of trying to synthesize what we have learned in incentive theory. We are grateful also to Cécile Aubert, Gary Biglaiser, Xavier Carpentier, Pierre Dubois, Yolande Hiriart, Wu Kai, Fahad Khalil, Claudio Mezzetti, Jérôme Pouyet, and Stéphane Straub, who have offered comments on the first draft and to Marie-Pierre Boé who cheerfully and repeatedly revised our chapters.

The Theory of Incentives

Contents

Introduction

It is surprising to find that Schumpeter (1954) does not mention the word "incentives" in his monumental history of economic thought. Today, for many economists, economics is to a large extent a matter of incentives: incentives to work hard, to produce good quality products, to study, to invest, to save, etc. How to design institutions that provide good incentives for economic agents has become a central question of economics.

Maybe Schumpeter's omission arose because, when he was writing, economics was mostly concerned with understanding the theory of value in large economies. For that purpose, neoclassical economics in particular postulates rational individual behavior in the market. In a perfectly competitive market, this assumption translates into profit maximization for firms' owners, which implies cost minimization. In other 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, such as workers, supervisors, and managers, with profit maximization. When economists began to look more carefully at the firm, either in agricultural or managerial economics, incentives became the central focus of their analysis. Indeed, for various reasons, the owner of the firm must delegate several tasks to the members of the firm. This necessity raises the problem of managing information flows within the firm. The problem of managing information flows was the first research topic for economists, once they mastered behavior under uncertainty, thanks to Von Neumann and Morgenstern (1944). This line of research culminated in the theory of teams (Marschak and Radner [1972]), which 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 on by Marschak (1955) and also by Arrow when he observes that

> by definition the agent has been selected for his specialized knowledge and the principal can never hope to completely check the agent's performance (1963a).

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 problem is the essence of incentive questions. If the agent had a different objective function but no private information, the principal could propose a contract that 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, incentive issues would disappear.

Conflicting objectives and decentralized information are thus the two basic ingredients of incentive theory. The essential paradigm for the analysis of market behavior by economists is one where economic agents pursue, at least to some extent, their private interests. What is proposed by incentive theory is that this major assumption be maintained in the analysis of organizations, small markets, and any other kind of collective decision-making. This paradigm has its own limits. Social behavior, particularly in small groups, is more complex, and norms of behavior that are culturally inculcated or developed over time 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 incen-

tives 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 readers not only to know more about incentive theory, but also to apply this indispensable tool when thinking about society.

The starting point of incentive theory corresponds to the problem of delegating 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 knowledge about his cost or valuation that is ignored by the principal, the case of *adverse selection* or *hidden knowledge*. Incentive theory considers 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 that has been raised in the literature is the case of *nonverifiability*, which occurs when the principal and the agent share ex post the same information but no third party and, in particular, no court of law can observe this information. One can study to what extent the nonverifiability of information is also problematic for contractual design.

We will discover that, in general, these informational problems prevent society from achieving the first-best allocation of resources that could be possible in a world where all information would be 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 modelling and analyzing these types of costs and providing a good understanding of the limits set by these on the allocation of resources. This work shows that the design of proper institutions for successful economic activity is more complex than one could have thought a priori. This line of research also provides a whole set of insights on how to begin to take into account agents' responses to the incentives provided by institutions.

As the next chapter will illustrate, a brief look at the history of economic thought shows that incentive theory was pervasive in many areas of economics, even though it was not central to economic thinking. Before describing how we will present this theory, it may be worth mentioning how the major achievement of economics, namely the general equilibrium theory (GE), met incentives.

General equilibrium theory was capable of producing powerful generalizations and able to deal with uncertainty, time, externalities, and extending the

---

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

validity of the *invisible hand* as long as the appropriate competitive markets could be set up.[2] 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 be satisfactorily imbedded in a proper generalization of the Arrow-Debreu theory. The problems encountered were so serious that a whole generation of general equilibrium theorists momentarily gave up the grandiose framework of GE to reconsider the problem of exchange under asymmetric information in its simplest form, i.e., between two traders. In a sense, the theorists went back to basics. They joined another group trained in game theory and in the theory of organizations, and together they built the theory of incentives, which we take as encompassing contract theory and mechanism design.

We will present incentive theory in three progressive steps. This book is the first step; in it we consider the principal-agent model where the principal delegates an action to a single agent through the 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 that are a topic for game theory.[3] Second, we assume the availability of a benevolent court of law that is able to enforce the contract and impose penalties if one of the contractual partners adopts a behavior that deviates from the one specified in the contract.[4]

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 cost. 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 book, 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.[5] This simple focus will turn out

---

[2] See Mas-Colell, Whinston and Green (1995) for a recent textbook exposition.

[3] See, for example, Osborne and Rubinstein (1994) and Muthoo (1999).

[4] 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 is minimal in what follows. However, judges must be given proper incentives to enforce contracts. We may rely here on the idea that in repeated relationships the desire to maintain their reputation will provide the appropriate incentives. This latter 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 in repeated relationships even in the absence of any contract, with the appropriate cooperative behavior being self-enforcing.

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

to be 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 objective 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.

Our next book 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, maintaining 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 that may 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 one agent by another are all mechanisms designed by the principal with this purpose in mind. Third, the mere attempt by the principal to use competition among agents may also trigger their collusion against the principal. The principal must now worry not only about individual incentives, but also about group incentives in a multiagent organization.

Our third book will be the final step of the analysis and will study the implications of various imperfections in the design of contracts: informed principal, limited commitment, renegotiation, implicit incentives, imperfect coordination among various principals, and incomplete contracting due to the nonverifiability of a parameter relevant for assessing the value of trade. The dynamics of some of these imperfect contractual relationships call for the extensive use of another equilibrium concept, the perfect Bayesian equilibrium. Equipped with this tool, we will be better able to describe the allocation of resources resulting from such imperfect contractual relationships.

In this book we proceed as follows. Chapter 1 gives a brief account of the history of thought concerning incentive theory. It shows that incentives questions have been present in many areas of economics over the last two centuries, even though it is only recently that their importance has been recognized and that economists have undertaken their systematic treatment. Chapter 2 presents the