Order Number 9405894

Information systems and service management: Information asymmetry and objective conflicts

Wang, Tswen-Gwo, Ph.D.

The University of Rochester, 1993

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

Information Systems and Service Management: Information Asymmetry and Objective Conflicts

By

Tswen-Gwo Wang

Submitted in Partial Fulfillment
of the
Requirement of the Degree of
DOCTOR OF PHILOSOPHY

Supervised by:

Professor Terence M. Barron

William E. Simon Graduate School of Business Administration
University of Rochester
Rochester, New York 14627

1993

CURRICULUM VITAE

The author, Tswen-Gwo Wang, was born on July 23, 1960 in Taichung, Taiwan, Republic of China. He obtained the Bachelor's degree in Business Administration in Accounting from the Tunghai University in 1982. After two years of military service and one year of working as an accountant at Asia Cement Co. in Taiwan, he attended the University of Washington (Seattle) between 1985 and 1987 and received the Master Business Administration degree. He entered the doctoral program at the William E. Simon Graduate School of Business Administration, University of Rochester in the summer of 1987, specializing in Computers and Information Systems, and Operations Research. He pursued his research in Information Systems Economics under the supervision of Professor Terence M. Barron. While at the Simon School, he also received the Master of Science degree in Business Administration with a specialization in Operations Research and Management in 1990.

ACKNOWLEDGEMENTS

I wish to sincerely thank the members of my dissertation committee, Professors Terence M. Barron, William B. Richmond, and Abraham Seidmann for their assistance in preparing this dissertation and support during my stay at the Simon School. This dissertation has also benefited from comments by Professors Eugene Kandel and Jinhong Xie.

I am especially indebted to my thesis advisor Professor Terence M. Barron. Professor Barron patiently read through innumerable drafts of my work, his insightful criticism and comments made this dissertation possible. Working with him is also one of the most rewarding experiences. I am very grateful to Professor Seidmann for his encouragement and support, both inside and outside of the school. I have also benefited greatly from our discussion on various research issues. Financial support from William E. Simon School fellowships is gratefully acknowledged. Finally, I am indebted to my parents, sisters, and wife, Li-Mei, for their love, encouragement, and patience. Without these this work would not have been possible.

ABSTRACT

The problem of central management's ability to control an information system (IS) department whose manager may have objectives that differ from those of the organization as a whole, and has informational advantages with regard to information system costs, is studied. A queueing model of a computing system is combined with a mechanism design approach for modeling the interaction between a firm's general management and its IS manager. Results are given, characterizing the optimal mechanisms for the above model for both cost center and profit center organizations for the IS department.

For cases with unlimited communication and no monitoring, by appealing to the revelation principle, my results show that the expected organizational net value is higher when the IS department is evaluated as a cost center governed by the optimal truth-revelation mechanism than when the IS department is evaluated as a profit center. This provides a general guideline on how the IS department should be controlled and evaluated. Our results may therefore help explain why organizing IS departments as a cost center rather than a profit center is more prevalent in practice.

By extending the basic model, the effects of three model variations are examined:

1. A finite feasible set of system choices;
2. A limited communication channel;
3. An imperfect, noiseless monitoring system about the IS department's reported costs.

Based on these results, a variety of managerial implications for evaluating and organizing IS departments are presented.

Table of Contents

List of Tables

List of Figures

List of Symbols

$\mathcal{R}_+$	the set of nonnegative real numbers
λ	the mean job arrival rate of the system
μ	the system's effective capacity
θ	the parameter that parameterizes the IS department's costs
Θ	the support of θ for the incomplete information case
$F(\theta)$	the cumulative probability distribution of θ
$V(\lambda)$	the users' aggregate value function
$W(\lambda, \mu)$	the users' delay cost function
$C(\mu, \theta)$	the information system cost
p	the price that the IS department charges the users
T	the transfer from the central management to the IS department
ξ	the index of the strength of the IS manager's incentive problem
$S(\cdot)$	the IS department's excess budget allocation
$U(\cdot)$	the IS manager's indirect utility function
π	the IS department's profit when it is organized as a profit center
$NV(\cdot)$	the organizational net value
$H(\cdot)$	the virtual organizational net value
$\mathcal{K}$	the set of feasible systems
$\mathcal{P}$	a partition of Θ
$\mathcal{M}$	the message set of the limited communication system
$COM(M)$	a limited communication system with M messages
$MON(\delta)$	a parameter-bound monitoring system with bound δ
$MON(s)$	a space-partition monitoring system with s intervals

Chapter 1

Introduction

1.1 Literature Review and Motivation

Every organization has limited resources, and these resources typically must be shared by the members of the organization. When the demand for these resources is greater than the supply, the problem of rationing arises. This problem is typically faced by an organization in managing its internal services. The primary purpose of rationing internal services is to control the services demanded through pricing or cost allocation so as to maximize the net value of the organization. A rationing scheme, then, is a mechanism for allocating resources and for providing information about the appropriate scale of the resources to be supplied. Rationing services is not limited to rationing internal services of an organization; studies dealing with rationing issues in the public sectors such as public utility pricing (Sharkey [96]), airport runway construction cost allocations and landing fees (Balinski and Sand [6]; Carlin and Park [14]; Littlechild and Thompson [67]; Oum and Zang [92]), cost allocation in telecommunication networks (Curien [19]; Sharkey [97]), land use (Oron et al. [91]; Solow [102]), and public transportation and highway tolls (Case and Lave [17]; Keeler and Small [55]; Mohring [79, 80]; Mohring and Harwitz [81]) abound in public economics literature.

To derive an effective mechanism for rationing an internal service, the mechanism designer must have sufficient information about the value and the cost of the services to the organization as a whole. The costs may appear as externalities like congestion delays as the members of an organization compete for the services. It is well-known that allowing the users to access a congestion-prone service system free of charges will result in an overcongested system (Bell and Stidham [11]; Mendelson [72]). With the presence of externalities such as congestion delays, using prices to internalize these externalities is common in both pubic and private organizations. In this dissertation, I will focus in part on the control of an organization's computing services through pricing.

Relevant Costs. Rationing internal services can be very difficult because there are many relevant, costly-to-measure opportunity costs (or externalities) associated with resource utilization. One example of such opportunity costs associated with regulating a congestion-prone system is queueing effects (Mendelson [72]); as an internal service system such as a computer system becomes more congested, the quality of service degrades because of a longer queue. Due to these externalities and because the value of a job submitted for services is known only to the users, and thus is hard to measure, the problems of measuring and modeling the users' valuation as well as delay costs must be resolved before an appropriate rationing scheme can be developed. Although information extraction can be endogenous to a control mechanism, giving it the dual purposes of extracting information and making decisions, the appropriate cost concept for the decisions being made must be determined with care.

A number of studies have focused on estimating various forms of the delay cost function in public sectors (e.g., Dewees [26], Keeler and Small [55], Vickery [107], and Walters [108] for highway congestion; Carlin and Park [14] and Morrison [82] for airport congestion). The work by Mendelson [72] was an important advance in the study of IS (information service or system) management problems due to its rigorous modeling of delay costs together with service capacity decisions. His model has been generalized in a number of subsequent papers, such as Dewan and Mendelson [25] and Whang [112]. Since externalities can be created in the form of congestion when the members of an organization compete for computing services, the problem of controlling such externalities is essential for maximizing an organization's net value.[1]

[1]The organizational net value of computing services is defined as the users' total gross value of the

Marginal Cost Pricing. Economists have long recognized that the social marginal cost pricing scheme must take into account external congestion costs in order to achieve the "first best" solution to the pricing of congestion-prone systems. The use of a pricing mechanism to regulate congested systems with a fixed capacity was first studied by Naor [88], and subsequently generalized by Dolan [28], Edelson and Hildebrand [31], Knudsen [59], Lippman and Stidham [66], Mendelson and Yechiali [75], and Yechiali [117], among others. These studies suggest that users should pay for the opportunity costs of the other users that result from a more congested system in order to achieve social welfare maximization. That is, in order to achieve system-wide optimality, the price of services must be able to internalize fully one particular user's external costs imposed on others.

For rationing computing services with perfect information about the users' aggregate value function, Mendelson [72] shows that charging a price equal to the delay costs that the marginal job inflicts on all other jobs in the system maximizes the organization's net value. For a short-run problem, i.e., when the system's capacity is fixed, there are many alternative mechanisms that can achieve the maximum organizational net value (Whang [112]).

For a long-run problem, the appropriate scale of the operation must be determined as well. Thus, capacity determination and capacity pricing become relevant. With queueing effects, the effect of an increased capacity is more complicated. First, for a fixed job arrival rate, an increase in capacity will reduce the delay costs incurred by all jobs, thus reducing externalities. Second, due to this reduction in delay costs, a higher arrival rate may be induced, thus increasing externalities. Depending on the functional form of the users' delay costs and the charges, a system can become more or less congested after a system's capacity has been altered. It is well-known that the short-run marginal cost evaluated at that capacity must equal the long run marginal cost in order to set the optimal capacity for a desired job arrival rate (see, e.g., Hirshleifer [50, 51]). Consequently, the optimal "capacity" charges may be higher or lower than the marginal

jobs being served minus the sum of the delay costs incurred by all jobs and the costs of providing the services.

capacity cost, and, due to queueing effects, the optimal pricing policy in general will not lead to an exactly balanced budget. In fact, this phenomenon is common, since most queueing systems exhibit increasing returns to scale. Within an organization, an unbalanced budget should not be a severe problem, since the trade relationship between the users and the IS department is not closed; the central management can always serve as a government to collect "taxes" or provide "subsidies."

With incomplete information about the users' aggregate value function, the incentive problem accompanying pricing schemes that rely on information supplied by users is obvious. Since the optimal capacity must depend on the users' valuation of the services, whenever the system manager lacks complete information, the users will have incentives to exaggerate their valuation so as to induce the system manager to acquire as large a system as possible. This problem may be countered by Groves mechanism (Groves [42, 43]; Groves and Loeb [44]). Under very mild conditions, Groves mechanism can yield an outcome in which all potential users find that revealing their true demand of the services is a dominant strategy. Although Groves mechanism in general does not yield a tidy allocation of costs for the IS department, it can always at least fully recover the capacity cost if the gross value of services is sufficiently high. Again, a mechanism that does not yield an exactly balanced budget should not be a big drawback for rationing an organization's internal services. Since a social marginal cost pricing scheme is optimal in rationing congestion-prone systems and does not always lead to a tidy allocation of costs, why in practice do firms use cost allocation schemes instead of marginal cost pricing schemes?

Common Cost Allocations. Allocating common costs is perhaps one of the most controversial issues in the accounting literature. Common costs arise when production costs are defined on a single intermediate product or service used by two or more users (Biddle and Steinberg [12]). Cost allocation schemes can affect an organization's value because in many circumstances "cost allocations, managerial behavior, and structure of the organization, including the incentive facing the managers, are inextricably linked" (Zimmerman [119]). Consequently, in practice cost allocations can be used, at least in part, to solve certain organizational control problems. Like marginal cost pricing schemes, cost allocations can be used to ration an organization's internal services.[2] Since opportunity costs are costly to measure in general, Zimmerman [119] has pointed out from the positive perspective that using allocated costs to proxy opportunity costs may be the optimum pricing scheme for internal resource utilization, given the costs of implementing and operating alternative systems.

However, except under very restrictive conditions, the optimal cost allocation scheme seldom leads to a tidy allocation of costs. One special case in which the system-wide efficient cost allocation scheme yields a tidy allocation of costs is studied by Whang [111].[3] Two assumptions cause the full cost allocation scheme to mimic the congestion external costs. First, the capacity cost function exhibits constant returns to scale. This assumption is reasonable as a first-order approximation since, if we ignore the administrative costs, a linear capacity cost function in terms of millions of instructions per second (MIPS) cannot be refuted empirically (Barron [10]; Mendelson [73]). Second, the users' delay function is homogeneous of degree zero. This assumption is more problematic because most queueing systems do not have this property. This makes the result not very useful even for controlling the most common queueing systems, such as $M/M/1$ and $M/G/1$. For example, if the system can be characterized as an $M/M/1$ queueing system with linear delay cost (homogeneous of degree minus one) and capacity cost

[2]Research in accounting on cost allocations largely focuses on the "fairness" of cost allocation methods; see, e.g., Biddle and Steinberg [12] and Young [118].

[3]This result is not new; it was identified by Mohring and Harwitz [81] and Mohring [79, 80] as they studied transportation systems. An interesting result identified by Whang [111] is that full cost allocation can also induce users to reveal their demand truthfully, thus achieving the ex post full efficient *capacity* decision. In order for this scheme to work, the IS manager must be able to make a credible commitment not to switch to marginal cost pricing even if capacity is set erroneously. In reality, however, capacity must be purchased before the *actual* service demanded can be realized. If the IS manager seeks to maximize organizational net value and we allow the users to form a tacit coalition, as in Whang [111], the users will still have incentives to exaggerate their demand, since a full cost allocation scheme may not be ex post implementable if the capacity was set too large. Consequently, the IS manager will have to switch to the marginal cost pricing scheme, and the full cost allocation scheme again fails to induce the users' truthful demand revelation.

functions, the optimal price equals the constant marginal capacity cost (Dewan and Mendelson [25]). Since the mean number of jobs served per unit of time must be less than the service rate, the optimal pricing scheme fails to recover the capacity cost. This is also true for an $M/G/1$ queueing system with a constant coefficient of variation, which is also homogeneous of degree minus one. Moreover, if an organization decides to allocate capacity costs partially by charging users the constant marginal capacity cost, the users will have incentives to exaggerate their valuation so as to minimize the service delays and thus the price ex post. As a result, this cost allocation scheme fails to solve the information problem, and therefore Groves mechanism may be more robust for inducing the users' truthful demand revelation.

According to Zimmerman [119], allocating the resource manager's expenditures to the users can create incentives for the users to monitor the resource manager's decisions, thus reducing the resource manager's "overconsumption-of-perquisites" problem. Of course, if the users have access to the resource manager's local information, cost allocations may turn out to be the least costly monitoring system. However, when the users do not have access to the resource manager's local information or the knowledge required to justify the IS manager's expenditures, relying upon the users as monitors will not be effective. Thus, when deriving an IS resource control mechanism, the IS manager's potential incentive problems must be addressed.

IS Department's Information and Incentive Issues. With perfect information, the central management may easily calculate the external costs to charge and the optimal capacity to set. Without perfect information, there are at least two kinds of incentive constraints that limit the ability of the central management to control the IS manager's investment in IS resources. First, when the IS manager has unverifiable private information that is either not available to the central management or simply too costly for the center to obtain by itself, then the manager cannot be compelled to reveal that information unless he or she is given the correct incentives. Second, when the IS manager controls some delegated decision variables whose outcomes cannot be effectively monitored by the central management, the manager cannot be directed to choose any particular action desired by the central management. These problems are natural consequences of specialization. In existing research on pricing policy and budget allocation for computing services, the IS manager is assumed to behave as a team member with the central management or is ignored (Dewan and Mendelson [25]; Mendelson [72]; Mendelson and Whang [76]; Whang [110, 111]).

Of course, in the absence of incentive problems, there is no motivation for the central management to exercise its control over the pricing and capacity decisions if the IS department possesses superior information about the demand and cost of the computing services. Thus, from the central management's standpoint, these information constraints do not really impose any loss of efficiency. But with conflicting objectives, the private information possessed by the IS department can introduce some serious problems. There are many incentive problems associated with IS professionals; *asymmetric cost*, *professionalism* (or *professional syndrome*), and *empire-building* are well-known (Mendelson [74]). In this dissertation, I focus on the IS manager's professionalism problem.

IS professionals are strongly motivated to learn state-of-the-art information technologies, since knowledge of these technologies is a major factor in determining their job market value (see, e.g., Couger [18]); this behavior is commonly called "professionalism." Professionalism may motivate IS staff to acquire software and hardware that are not cost-justified in terms of organizational net-value maximization if such decisions are left to them. Thus, a natural tendency may arise for the IS manager to desire a very high quality department even though a somewhat lower quality department would provide almost the same performance at significantly lower costs (Kaplan and Atkinson [54], p. 531). The professionalism problem is similar to the "overconsumption-of-perquisites" problem raised by Zimmerman [119]; the IS department's decision-makers tend to overinvest in state-of-the-art information technologies. This behavioral assumption is reasonable since the IS managers strive to maximize utility rather than wealth, and consuming perquisites increase their utility (Biddle and Steinberg [12]). Consequently, the ability of the central management to control and motivate IS professionals is further impaired if monetary rewards (at least in the short-run) may not be the sole factor driving their behavior and the consumption of perquisites is difficult to monitor.

Moreover, the IS profession is highly specialized. Within an organization, the IS staff can be viewed as the delegated experts specializing in acquiring knowledge about computer technologies and the system operating environment. They not only have the advantages of their expertise on general computer technologies but also have more information about the characteristics of the current system as well as the users' demand. As Demsetz ([22], p. 172) has pointed out: "Although knowledge can be learned more effectively in specialized fashion, its use to achieve high living standards requires that a specialist somehow use the knowledge of other specialists. This cannot be done only by *learning* what others know, for that would undermine gains from specialized learning. It cannot be done only by *purchasing* information in the form of facts, for in many cases the theory that links facts must be mastered if facts are to be put to work" (emphases are original). Determining the IS department's future investment or budget allocation therefore requires the judgment of and the directions given by the informed IS professionals, especially the IS manager. As a result of this informational decentralization and asymmetry, the IS manager might misrepresent the information that he or she has in order to use organizational resources to facilitate personal interest-seeking activities, e.g., overinvestment in technologically advanced staff and/or computing technologies. This "opportunism" behavioral assumption for economic agents is typical in the literature of institutional economics (Williamson [116]). Thus if the objectives of the IS manager do not totally coincide with those of the central management, this gives the IS manager an obvious opportunity to misrepresent her private information. This type of "informational rent" that the IS manager can capture might be viewed as "returns to specialization or expertise."

Furthermore, the IS department itself is often a very complex organization with many different activities, e.g., systems development, hardware and software maintenance, etc. It is very hard to judge the effectiveness and efficiency of the IS department without required expertise. Traditional cost or management accounting provides few useful tools or guidelines to control or evaluate the performance of IS departments. Personal communication with various personnel of large corporations has revealed that, even if the IS department is called a cost center, it may in fact be functioning as a discretionary expense center.[4] Kaplan and Atkinson [54] note: "Given the difficulty of measuring the efficiency of discretionary centers, a natural tendency may arise for their managers to desire a very high quality department even though a somewhat lower quality department would provide almost the same service at significantly lower costs. Accentuating this tendency, the white-collar professionals who typically staff these centers prefer to have the best people in their discipline associated with them so that they can take pride in the quality of their department" (p. 531–532). Thus, given the IS manager's expertise and potential incentive problems, it is unlikely that the manager of an IS department will attempt to maximize the organization's net value. It is then more appropriate to view the budget or investment funding of the IS department as the outcome of an internal bargaining game between the central management and the IS manager.

Organization of IS Departments. Obviously, the central management should provide some incentive scheme (reward or punishment) which aligns the IS department's objectives more closely with the central management's. If there were no uncertainty or informational asymmetry, even with conflicting objectives, designing an efficient incentive scheme would be a trivial task for the central management, since all relevant performance criteria could be specified in the contract. (However, the contracting process itself might become very costly in order to specify all the relevant factors and contingencies.) Yet, even in a very small organization, it is usually too costly for the owner to acquire all the information required to make efficient decisions, e.g., the communication costs, the owner's information processing and optimization costs, etc. This provides an obvious reason for the central management to decentralize decisions given that some constraints can be imposed on the behaviour of the IS department. Nevertheless, there

[4]Discretionary expense centers are used for organizational units that produce outputs that are not measurable in financial terms or for units where no strong relation exists between resources expended and results achieved. When great information asymmetry is likely to exist between a unit's local manager and central management, it is more appropriate to organize that local unit as a discretionary expense center, and the budget allocation usually is a result of internal bargaining. For a detailed discussion on discretionary expense centers, see Kaplan and Atkinson [54].

will be some efficiency loss (i.e., the agency cost) if the central management is not able to monitor the behavior of the agent effectively. Uncertainty about the IS department's productivity and the cost function of computing services further impairs the central management's ability to motivate and control the IS department. The moral hazard induced by asymmetric information is a well-recognized phenomenon in the literature of agency and contracting theory.

There are two much-studied forms of organizing and evaluating information services units: the cost center and the profit center. Cost centers are effective when the output can be defined and measured well and the required inputs per unit of output can be specified (Kaplan and Atkinson [54]). Some outputs of an information system in terms of system performance measures are not particularly difficult to measure, e.g., the average number of jobs processed per unit of time and the average response time. However, as discussed earlier, information systems are a complex bundle of hardware, software, operating personnel, and users, so the inputs required to achieve a specific level of performance are hard to measure. The relationship between inputs and outputs is further blurred by such qualitative requirements as the degree of user friendliness and graphical interface, etc. Without the required knowledge and information about computer technologies and the local operating environment, it is difficult for the central management to judge the standard costs appropriate to attaining a specific level of system performance. If the central management realizes the IS manager's incentive problems, it will not merely accept whatever standard cost or budget allocation is proposed by the IS manager; some form of bargaining game may be expected during the budgeting process. Within an organization it is natural to assume that the central management possesses all the bargaining power and thereby has the ability to prescribe any budget allocation rule that it sees fit.

When the central management does not possess all relevant information to effectively determine and control the IS department's actions, there are several questions that need to be addressed: How can the central management motivate the IS manager to reveal her private information truthfully? How can the central management evaluate the appropriate amount of "subsidy" or "taxation" if it does not have all the relevant information about the IS department's cost? Should the central management admit whatever surplus or deficit shows on the IS department's budget account? How is the central management able to judge that the system is in a proper state and the IS department uses the resources properly? To resolve these issues, the way that the IS department is evaluated must be endogenous to an IS resource control mechanism.

According to the data gathered by McGee [69], a little over 8% of the companies in his sample use a profit center to manage their computing resources. A profit center is one form of decentralized mechanism where the capacity and pricing decisions are delegated to the IS department.[5] When an IS department is organized as a profit center, its performance is evaluated based on the profit it generates. In the absence of outside markets for the services, the IS manager will have incentives to behave like a monopolist; the pricing decision that maximizes profit will not maximize the organizational net value in general, even though organizing the IS department as a profit center encourages efficient use of the organization's resources. This problem stems from the distortion of the IS department's valuation of the users' jobs, not the costs. Consequently, it is never optimal for an organization to operate its IS department as a profit center due to the monopolistic pricing problem (see Dewan and Mendelson [25]; Mendelson [72]). These facts also illustrate how the managerial behavior and the structure of the organization are inextricably linked.

This dissertation complements previous research in an important way by formally modeling, using a mechanism design approach, the information asymmetry and objective conflicts that can be expected to exist between the IS manager and the central management. Since the issues related to pricing computing resources with both perfect

[5] Allen [1] advocates treating IS departments as profit centers, but he primarily emphasizes to run an IS department "with a flexible budget and a systematic way to price its services." He also stresses that the revenue of a profit center must equal or exceed cost. His "profit center" thus differs from the usual profit center, in which maximization of profit is stressed. Moreover, as we discussed earlier, after accounting for those hard-to-measure opportunity costs, the revenue generated by the optimal pricing scheme does not equal or exceed cost in general. More importantly, he fails to recognize that "organizational slack" may still appear as the form of profits if the incentive problems of IS managers were not appropriately addressed.

and imperfect information have been studied extensively, I treat the users' valuation of IS services and delay costs as common knowledge and focus on the supply side of IS resource management. Using the meta model, introduced in the next section, I compare the performance of an IS department organized as a cost center governed by a centralized mechanism with its performance when organized as a decentralized profit center, with several variations in environment.

1.2 Meta Model

Although the basic game-theoretic approach employed in this research is familiar from the mechanism design literature (Fudenberg and Tirole [32]; Myerson [86]), I add several features significant in the IS setting to make the work more realistic. In order to clarify the problems and their relationships to one another, I introduce the meta model shown in Figure 1.1 and discussed below.

Figure 1.1 shows the seven basic components of the meta model used in my analyses. The two interacting parties, the central management and the IS department represented by its manager, and the control mechanism are standard in the principal-agent and mechanism design literature. The information system is represented by a queueing system with a single service center, which is characterized by the mean job arrival rate and system capacity, (λ, μ). The restriction on the set of feasible systems and the presence of a potentially limited communication system and monitoring system depart from the traditional queueing-based research on controlling computer systems.

Feasible Systems. In most studies on setting capacity for queueing systems, the set of feasible systems is usually assumed to be continuous (e.g., Dewan and Mendelson [25]; Stidham [104]; Whang [111]). However, in a real-world business environment, a firm has a finite number of systems, often just a few, from which to choose. When the set of feasible systems is finite, even though the central management's prior beliefs about the possible realizations of the IS department's cost is distributed continuously, the central management's ability to induce the IS manager's truth-revelation is further limited since the system must be the same over certain ranges of parameter values. In other words,

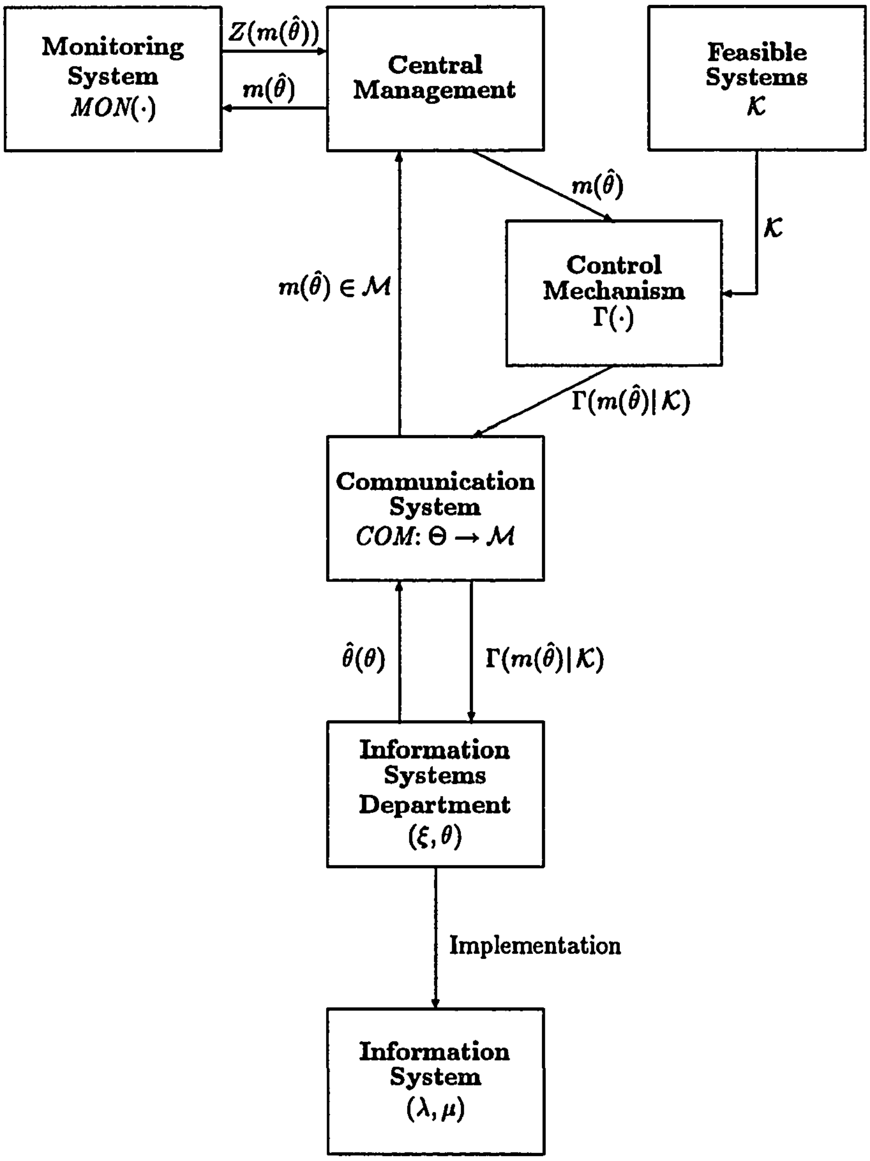


FIGURE 1.1: META MODEL.

some degree of pooling or bunching over the "types" of the IS department is inevitable.

The problem associated with a finite set of feasible systems can also arise when the organization's computing system can be characterized as a multi-server queueing system with a fixed capacity for each server. In this situation, determining the capacity is equivalent to determining the number of servers for the system. Consequently, my analysis can be easily extended to study systems with multiple processors.[6]

Communication Systems. Within an organization, there are many sources which can limit communication. First, information is quite often channeled through budgets and divisional profit reports, which typically condense a large amount of information into a summarized form. It has been shown at least theoretically that the way the accounting numbers are aggregated can affect the behavior of members within an organization (Demski and Sappington [23]).

Furthermore, the language used to communicate is imprecise in general; the same message may have different meanings for different receivers. This is particularly true if interpreting a message correctly requires expertise. Since the members of an organization have a limited information processing capacity, a message can be technically received but ignored or interpreted carelessly because the intended receiver is overloaded by messages. This assumption is common in behavioral and economic theories of organizations (Cyert and March [20]; Simon [100]; Williamson [116]). Full communication may thus be too costly even when possible. For example, it may require a significant amount of time and effort for the IS manager to prepare a detailed report about her department's local operating environment. It may take even more time and effort for the central management to understand the IS manager's report. A similar condition holds when the central management tries to design a mechanism and communicate it to the IS manager.

Guesnerie and Laffont [45] show that, with a single agent, if a decision function is truthfully implementable by a "direct" mechanism, it can be equivalently implemented

[6]Although in this dissertation I focus on systems with a single service center, the applicability of my analysis to multi-server systems will be illustrated by an example in Chapter 4.

by an "indirect" mechanism, in which the agent chooses a decision (or a particular transfer from a full schedule of transfers offered by the central management) rather than announcing her private information. However, an indirect mechanism does not save much communication between the central management and the IS manager because the central management still must explain the mechanism to the IS manager and demonstrate that she is bound to it.

Conceivably, if the central management has the ability to design a complete schedule of decision rules and communicate it effectively, it should have all the required expertise to justify and evaluate the performance of its IS department. Without the required expertise, the mechanism that the central management can design will be constrained. To approximate this "limited expertise" phenomenon when the central management centralizes the IS-related decisions, I assume it can only design a "crude" mechanism governed by a "limited communication system." This assumption departs from the traditional principal-agent and mechanism design models. Of course, limited communication would have no effect if all IS-related decisions are decentralized. As a result, the performance of a profit center may not be replicated by a centralized mechanism given the limitations that I impose on the central management. Details of the limited communication model will be introduced in Chapter 5.

Monitoring Systems. In most existing literature studying centralized organizational decision-making mechanisms, the range over which an economic agent can misrepresent private information is typically bounded only by the mechanism designer's prior beliefs. This assumption will be relaxed in Chapter 6. When the IS manager has incentives to misrepresent her private information, I investigate the central management's ability to control its IS department in an environment where the central management possesses an imperfect but noiseless monitoring (partially verifiable) information system. The monitoring system is imperfect because the central management cannot detect and verify the IS manager's false report as long as it is not "too big" a lie. A monitoring system is said to be noiseless if the central management can detect and verify the IS manager's false report with certainty whenever the difference between a false report and truth is beyond a certain limit specified by the monitoring system. The exact meaning of these statements will become transparent as I present my formal model in Chapter 6.

Discussion. The information system is characterized by a queueing system with job arrival rate and system capacity $(\lambda, \mu) \in \mathcal{R}_+ \times \mathcal{K}$. The IS department is characterized by two parameters, (ξ, θ), where ξ indexes the strength of the IS manager's incentive problem and θ summarizes the IS manager's private information about the IS department's costs. The IS manager is required to report her private information θ to the central management through the communication system, $COM : \Theta \rightarrow \mathcal{M}$, where $\mathcal{M}$ is the message set supported by the communication system. When the communication is unlimited, $\Theta = \mathcal{M}$, and when the communication limited, $\mathcal{M} \subset \Theta$. Depending on the realized θ and the control mechanism, $\Gamma(\cdot)$, prescribed by the central management, the IS manager reports $\hat{\theta}$ as the input to the communication system. If the communication is unlimited, the report received by the central management is $m(\hat{\theta}) = \hat{\theta}$; otherwise, $m(\hat{\theta}) \in \mathcal{M} \subset \Theta$.

Let $\mathcal{N}(m(\theta))$ be the range specified by the monitoring system, given that $m(\theta)$ is the "correct" message that the IS manager should report. Then, in the presence of a monitoring system, the IS manager's misrepresentation will not be detected if and only if $m(\hat{\theta}) \in \mathcal{N}(m(\theta))$. In Figure 1.1, $Z(m(\hat{\theta}))$ is a binary (non-stochastic) variable whose value is jointly determined by the correct message, $m(\theta)$, and the message received by the central management, $m(\hat{\theta})$. The value of $Z(m(\hat{\theta}))$ indicates whether or not $m(\hat{\theta}) \in \mathcal{N}(m(\theta))$. Notice that, since $\mathcal{N}(m(\theta))$ can vary with $m(\theta)$, it is important not to allow the central management to observe the entire set $\mathcal{N}(m(\theta))$; otherwise, the central management would be able to infer $m(\theta)$ itself. I also assume that the central management can impose a sufficient penalty to deter the IS manager's misrepresentation beyond $\mathcal{N}(m(\theta))$. Consequently, $\mathcal{N}(m(\theta))$ specifies the range that the IS manager can misrepresent. Of course, in the absence of a monitoring system, $m(\hat{\theta})$ is restricted only by the message set, $\mathcal{M}$. Finally, given $(m(\hat{\theta}), \mathcal{K})$ and the control mechanism $\Gamma(\cdot)$, the central management determines the decision variables (λ, μ) that the IS department should implement.

Since dealing with the meta model in its most general form is intractable, I focus on examining the effect of relaxing the assumptions imposed on individual components.

TABLE 1.1: SUMMARY OF THE CASES CONSIDERED IN THIS DISSERTATION. (× MEANS THE CASE IS NOT COVERED.)

	Continuous Systems			Finite Systems		
	Parameter-Bound	Space-Partition	No Monitoring	Parameter-Bound	Space-Partition	No Monitoring
L.C.	×	×	Chapter 5	×	×	×
U.C.	Chapter 6	Chapter 6	Chapter 3	×	×	Chapter 4

L.C.: Limited Communication

U.C.: Unlimited Communication

1.3 Organization of the Dissertation

The plan of this dissertation is as follows. In Chapter 2, I provide a review of the queueing model and mechanism design. The models analyzed and their corresponding chapters in this dissertation are given in Table 1.1. Chapter 3 analyzes the standard case with unlimited communication, a continuous set of feasible systems, and no monitoring. Three control mechanisms, the optimal incentive compatible mechanism, the naive mechanism, and the profit center, are analyzed and compared. In Chapter 4, I relax the assumption that the set of feasible systems is continuous. This chapter examines the pooling effect when the set of instruments (feasible systems) that the central management can use to discriminate the IS department's "types" is smaller the space of its "types." Chapter 5 examines the effect of limited communication between the central management and the IS department. Since unlimited communication is required for the revelation principle to be valid, in addition to deriving the optimal centralized mechanism, I demonstrate several cases where a profit center can outperform the optimal incentive compatible mechanism. In Chapter 6, two monitoring technologies, parameter-bound and space-partition, are analyzed. I show that a parameter-bound monitoring system has no value for inducing the IS manager's truth-revelation and can therefore cause the revelation principle to fail, while a space-partition monitoring system can improve the expected organizational net value (strictly) as long as the partition is not completely degenerate. Although I do not consider an environment with both limited communication and an imperfect monitoring system, extending my analysis to the case with limited communication and space-partition monitoring is straightforward. Combining limited communication with parameter-bound monitoring is left for future research. Finally, Chapter 7 contains some concluding remarks and discusses future research directions.

Chapter 2

Basic Framework of Analysis

2.1 Queueing Model

As in Dewan and Mendelson [25] (cf. Stidham [104]), I assume that the times between successive arrivals of jobs to the computer system are independent and identically distributed (i.i.d.) random variables with finite mean $\frac{1}{\lambda}$. The generic random variable representing interarrival times is denoted by $\tilde{t}$ and assumed to have the form: $\tilde{t} = \frac{\tilde{\tau}}{\lambda}$, where $\tilde{\tau}$ is a fixed random variable with unit mean, and the arrival rate λ is a scale parameter that depends on the value of the services to their users and on the corresponding user costs. The expected gross value of the information processing services to the organization per unit of time is given by the value function $V(\lambda)$, which aggregates the values of users' job requests corresponding to arrival rate λ. $V(\lambda)$ is assumed to be twice continuously differentiable, strictly concave over $\mathcal{R}_+$.

Following Stidham [104] (cf. Dewan and Mendelson [25]; Lippman [65]; Lippman and Stidham [66]), the value of jobs served can be considered to be a non-negative random variable Y with support in a convex subset of $[0, \infty)$ and with a continuous distribution function $\Phi = \Pr\{Y \leq x\}$. If all arriving jobs with values greater than x join the system, the resulting arrival rate is $\lambda = \Lambda\overline{\Phi}(x)$, where $\overline{\Phi}(x) = 1 - \Phi(x)$ and Λ is the maximal arrival rate. Conversely, when the arrival rate is λ, the marginal value x is equal to $\overline{\Phi}^{-1}(\lambda/\Lambda)$, where $\overline{\Phi}^{-1}$ is the inverse of $\overline{\Phi}$. Thus, $V(\lambda) = \Lambda \int_{\overline{\Phi}^{-1}(\lambda/\Lambda)}^{\infty} x d\Phi(x)$, and $V'(\lambda) = \overline{\Phi}^{-1}(\lambda/\Lambda)$. However, I work directly with the value function $V(\lambda)$ and ignore the constraint $\lambda \leq \Lambda$, as in Mendelson [72] and Dewan and Mendelson [25], i.e., $\Lambda = \infty$.

Jobs submitted to the system are served according to the First Come First Served (FCFS) queue discipline. Service times are assumed to be i.i.d. random variables denoted by $\tilde{x}$ and have the form: $\tilde{x} = \frac{\tilde{\chi}}{\mu}$, where $\tilde{\chi}$ is a random variable with unit mean and finite variance, and μ is a scale parameter representing the service rate or processing capacity of the computer system. In other words, the jobs submitted by the users are homogeneous in terms of their processing requirements. The distribution of service times is not restricted; it could be exponentially distributed or belong to some other family of general distributions.[1]

For a given pair $(\lambda, \mu) \in \mathcal{R}^2_+$, then, the expected delay cost per job is given by

$$W(\lambda, \mu) \stackrel{\text{def}}{=} E\{D(\tilde{W}(\lambda, \mu))\},$$

where $D(\cdot)$ is the users' delay cost function, which is assumed to be non-decreasing and convex, and $\tilde{W}(\lambda, \mu)$ is the job waiting time with a stationary distribution. Consequently, $W(\lambda, \mu)$ is increasing and convex in λ and decreasing and convex in μ (Dewan and Mendelson [25]; Stidham [104]). I assume $W(\lambda, \mu)$ is twice continuously differentiable with respect to λ and μ with cross-partials of constant sign: $W_{\lambda\mu} = W_{\mu\lambda} < 0$.[2]

The cost of capacity, $C(\mu, \theta)$, is assumed to be increasing in both μ and θ and twice continuously differentiable, where θ parameterizes the quality or the ability of the IS department. A higher θ represents a lower quality department, which will cost the organization more to operate at a given capacity.[3]

---

[1] This model can be easily generalized to multi-class cases with FCFS queue discipline as long as the process requirements are homogeneous (Dewan and Mendelson [25]). For a treatment on heterogeneous cases with priority queue disciplines, see Dolan [28], Mendelson and Whang [76], and Whang [110].

[2] See Widder [113] for sufficient conditions of differentiability and Heyman and Sobel [49] or Dewan and Mendelson [25] for a discussion of the generalization of Little's law. The condition that $W_{\lambda\mu} = W_{\mu\lambda} < 0$ is satisfied, for example, if the system is an $M/M/1$ or $M/G/1$ queue with linear user delay costs.

[3] In general, the production or cost function of an IS department is multidimensional, e.g., availability, quality, support services, etc., but the multidimensional case is intractable even for very simple functional

I assume that both the users and the central management have the same valuation for the expected gross value and delay cost; that is, the central management's objective is to maximize the aggregate net value provided by the information system. The users determine whether or not to have their jobs served based on the price and the expected queue length in steady state. In this research I focus on the case where system capacity is one of the decision variables.

With perfect information, the problem of determining the optimal price and capacity to maximize the organization's net value can be formulated as follows:

$$\max_{\lambda,\mu} V(\lambda) - \lambda W(\lambda,\mu) - C(\mu,\theta) \tag{2.1}$$

I call this program the *full-information* case and the corresponding solution the *full-information* solution. I assume that the Hessian matrix of (2.1) is negative definite for $(\lambda,\mu) \in \mathcal{R}^2_+$, and therefore (2.1) has a unique maximum. The solution of this unconstrained optimization problem comes from the following first-order conditions:

$$0 = V'(\lambda) - W(\lambda,\mu) - \lambda W_\lambda(\lambda,\mu) \tag{2.2}$$

$$0 = -\lambda W_\mu(\lambda,\mu) - C_\mu(\mu,\theta), \tag{2.3}$$

where the prime and subscript denote the first and partial derivatives, respectively. Equation (2.2) states that, at optimum, the marginal job's net value (net of its own delay cost), $V'(\lambda) - W(\lambda,\mu)$ should equal the marginal delay cost incurred by all other jobs in the system, $\lambda W_\lambda(\lambda,\mu)$ (i.e., the "external cost", following Mendelson's [72] terminology). That is, the jobs to be served should generate a net value at least as large as the "external cost" that the marginal job imposes on all other jobs existing in the system. In order to achieve system-wide optimality, therefore, the jobs should be charged a price equal to the "external cost."

Let λ^f and μ^f be the solutions of (2.2) and (2.3). The optimal price is:

$$p^f \stackrel{\text{def}}{=} V'(\lambda^f) - W(\lambda^f,\mu^f) \tag{2.4}$$

$$= \lambda^f W_\lambda(\lambda^f,\mu^f) \tag{2.5}$$

---

forms and left for future research. We may nevertheless consider θ as a summary statistic and μ as an aggregate performance measure of the system's effective capacity.

where (2.5) indicates that the optimal price should be equal to the "external cost," and (2.4) gives the *marginal* user's (job's) inverse demand function for the computer service. Hence, given any p and μ, the *induced* arrival rate, $\lambda(p,\mu)$ (i.e., the demand for the computing service), is determined by solving $p = V'(\lambda) - W(\lambda,\mu)$. Since with a fixed capacity, $\hat{\mu}$, the short-run marginal (external) cost is $\lambda^f W_\lambda(\lambda^f,\hat{\mu})$, we have by the envelope theorem the standard result: the optimal capacity is set where the short-run marginal cost equals the long-run marginal cost:

$$\lambda^f W_\lambda(\lambda^f,\hat{\mu})\Big|_{\hat{\mu}=\mu^f} = \lambda^f W_\lambda(\lambda^f,\mu^f).$$

Due to these queueing effects, the optimal "capacity" charge (or the long-run marginal cost) may not lead to a tidy allocation of the capacity costs. Then the optimal subsidy to cover the IS department's budget deficit or the optimal surplus that the IS department should show is equal to

$$T^f \stackrel{\text{def}}{=} C(\mu^f,\theta) - p^f \lambda^f. \tag{2.6}$$

T^f will not be useful in evaluating an IS department in general, since a balanced budget will almost never correspond to the result of a decision to maximize the organizational net value. Whether or not the price that maximizes the organizational net value will result in a balanced budget for the IS department depends on both the nature of the users' delay cost function and the capacity cost function. Specifically, let $W(\lambda,\mu)$ be the users' delay cost function when the system capacity is μ and the job arrival rate is λ. Then, depending on the homogeneity of $W(\lambda,\mu)$ and the cost function of the system capacity, the IS department's budget can be in balance, show a deficit, or show a surplus. As discussed in Chapter 1, Whang [111] shows that the IS department will have a balanced budget when the users' delay cost function is homogeneous of degree zero and the system capacity cost is linear. It is also straightforward to show that, with linear capacity cost, if the homogeneity of the users' delay cost function is negative (positive), the IS department will show a budget deficit (surplus) (see Dewan and Mendelson [25]). For instance, suppose that the computer system of an organization can be characterized as an $M/M/1$ queueing system with linear capacity and user delay costs; the IS department will show a budget deficit since the mean waiting time of an

$M/M/1$ system is $\frac{1}{\mu-\lambda}$, and the users' delay cost function is therefore homogeneous of degree minus one. One of the interesting features of $M/M/1$ queueing systems is that

$$\frac{1}{(\mu+\Delta\lambda)-(\lambda+\Delta\lambda)}=\frac{1}{\mu-\lambda}.$$

That is, for a given increase in the mean job arrival rate, $\Delta\lambda$, if we increase the capacity by the same amount, the negative externalities are completely neutralized, and therefore the mean job waiting time remains the same. So it should not be surprising that the optimal price equals the "marginal capacity" cost. Although, from (2.3) and by the implicit function theorem, $\frac{d\mu^f(\lambda)}{d\lambda} > 1$, the possible reduction in aggregate delay costs is countered by a higher mean job arrival rate. As a result, the optimal price remains equal to the marginal capacity cost, and the optimal pricing scheme will not fully recover the capacity cost if the capacity function is linear.

When the capacity cost is not linear, the above results fail to hold in general. For instance, we may have an $M/M/1$ queueing system with a quadratic capacity cost function and a linear user delay cost, but the IS department will show a surplus under the optimal pricing policy. In particular, as in Dewan and Mendelson [25] and Whang [111], if we replace $\lambda W(\lambda,\mu)$ in (2.1) by an aggregate users' delay cost function $G(\lambda,\mu)$, which is characterized by the degree of homogeneity, then we have the following simple results:

PROPOSITION 2.1 *Let $G(\lambda,\mu)$ denote the aggregate users' delay cost function, which is homogeneous of degree α, and let (λ^f,μ^f) denote the solution to the equations (2.2) and (2.3). Then*

1. *if $(\alpha-1)\,G(\lambda^f,\mu^f) > C(\mu^f,\theta)-\mu^f\,C_\mu(\mu^f,\theta)$, there is a budget surplus;*

2. *if $(\alpha-1)\,G(\lambda^f,\mu^f) = C(\mu^f,\theta)-\mu^f\,C_\mu(\mu^f,\theta)$, there is a balanced budget;*

3. *if $(\alpha-1)\,G(\lambda^f,\mu^f) < C(\mu^f,\theta)-\mu^f\,C_\mu(\mu^f,\theta)$, there is a budget deficit.*

PROOF. Note that the first order conditions imply that the optimal price equals $V'(\lambda^f)-\frac{G(\lambda^f,\mu^f)}{\lambda^f}$, since by Little's law $G=\lambda W(\lambda,\mu)$, and so the IS department's revenue is $\lambda^f\,V'(\lambda^f)-G(\lambda^f,\mu^f)$. By Euler's equation,

$$\lambda^f\,G_\lambda(\lambda^f,\mu^f)+\mu^f\,G_\mu(\lambda^f,\mu^f)=\alpha\,G(\lambda^f,\mu^f),$$

and from the first order conditions $V'(\lambda^f) = G_\lambda(\lambda^f, \mu^f)$ and $-G_\mu(\lambda^f, \mu^f) = C_\mu(\mu, \theta)$, the IS department's revenue then is

$$\lambda^f V'(\lambda^f) - G(\lambda^f, \mu^f) = (\alpha - 1) G(\lambda^f, \mu^f) + \mu^f C_\mu(\mu^f, \theta),$$

proving the theorem. ||

Note that, although the results in Theorem 2.1 are local, most of them hold globally as long as the marginal capacity cost function has constant sign. For instance, when $\alpha = 1$, the IS department's budget status resulting from the optimal pricing policy depends solely on the scale economies of the capacity cost and is independent of the location of the solutions, λ^f and μ^f. Specifically, whenever $\alpha = 1$, the IS department will show a budget surplus, a balanced budget, or a budget deficit as long as the system capacity exhibits diseconomies of scale, constant return to scale, or economies of scale. These properties hold globally. Similar conclusions can be derived for the cases where $\alpha \neq 1$ with two exceptions. First, when $\alpha > 1$ and the system capacity exhibits economies of scale, both $(\alpha - 1) G(\lambda^f, \mu^f)$ and $C(\mu^f, \theta) - \mu^f C_\mu(\mu^f, \theta)$ are positive. Therefore, depending on the specific forms of the gross value function and the capacity cost function, the IS department may show a deficit, a surplus, or even a balanced budget. Similarly, when $\alpha < 1$ and the system capacity cost exhibits diseconomies of scale, the IS department's proper budget status cannot be determined without specifying the gross value and capacity cost functions.

From Theorem 2.1, when $G(\lambda, \mu)$ is a homogeneous function, budget balancing corresponds to optimality regardless of the users' aggregate value function if and only if

$$\frac{C(\mu^f, \theta) - \mu^f C_\mu(\mu^f, \theta)}{G(\lambda^f, \mu^f)}$$

is a constant and equals $(\alpha - 1)$. Thus, as long as $\alpha \neq 1$, the linearity of the capacity cost function does not suffice to guarantee that a balanced budget corresponds to optimality, and a balanced budget is almost never an appropriate criterion for evaluating the IS department's performance. This demonstrates the central management's difficulty in effectively evaluating the IS department's performance based on its budget status.

By Theorem 2.1, it is obvious that, if $C(\mu, \theta)$ is linear in μ, we have Theorem 4 of Dewan and Mendelson [25]. Furthermore, if the system is $M/M/1$ (i.e., $\alpha = 0$), the IS department can still have a budget surplus if the marginal capacity cost is increasing. For example, suppose that $V(\lambda) = 10 \ln \lambda$, $G(\lambda, \mu) = \frac{\lambda}{\mu - \lambda}$ (i.e., the system is $M/M/1$ with unit delay cost), and $C(\mu, \theta) = \frac{1}{2}\mu^2$. Then $\lambda^f = 2.308$ and $\mu^f = 3.162$, and the IS department will have a budget surplus equal to 2.298. Thus, even though the mean job waiting time of an $M/M/1$ queueing system is homogeneous of degree minus one, a quadratic capacity cost function makes the marginal cost of a job upward-sloping (at least locally). Although constant return to scale in terms of million instructions per second (MIPS) (i.e., hardware costs) cannot be refuted empirically for computer systems (Barron [10]; Mendelson [73]), one must also take administrative and coordination costs and CPU overheads into account, so the total capacity costs incurred by an organization may well exhibit diseconomies of scale due to the complexity of managing a larger system.

2.2 Mechanism Design

When the IS department is organized as a cost center, I assume that it is governed by a centralized mechanism under which—based on the cost parameter θ reported by the IS manager—the central management determines the budget allocation and all the operating variables, such as prices and capacity. A game-theoretic approach is required to investigate whether or not the central management can design a set of decision rules to induce the IS manager's truth-reporting and at the same time eliminate or reduce the informational rents that the IS manager can command.

Let θ summarize the private information possessed by the IS department, which parameterizes the IS department's cost function, and let $f(\theta)$ be the central management's common knowledge prior beliefs concerning θ, where $f(\theta) > 0$ if and only if $\theta \in \Theta \subset \mathcal{R}_+$. Since the IS manager possesses some private information before choosing the strategies for playing the game, it is appropriate to view it as a game with incomplete information (Myerson [85]) and to seek a set of decision rules by appealing to the notion of *interim incentive efficiency* (Holmstrom and Myerson [53]). A set of decision rules is *(interim) incentive compatible* if and only if it induces the IS manager's truthful revelation of her private information. In other words, the IS manager maximizes her conditional expected utility or payoff by revealing her private information truthfully under a set of incentive compatible decision rules. A set of decision rules is ***interim incentive efficient*** if and only if it is incentive compatible and is not dominated by any other set of incentive compatible decision rules from the central management's standpoint.

Let D_1 be the feasible set of the IS manager's decisions, which by taking the mechanism approach is the message space of the IS manager's reporting regarding θ. Thus, given the space of the cost parameter and D_1, the IS manager's reporting strategy is a mapping: $\sigma_1 : \Theta \rightarrow D_1$. Let D_2 be the feasible set of decisions available to the central management, i.e., D_2 is the set of possible decisions related to the IS operations and budget allocation: λ, μ, and T.

Then we can view a mechanism as a process of generating decisions and allocating budgets (λ, μ, T) through a Bayesian game within which the central management and the IS manager act and interact by the following sequence of events:

1. The IS manager observes the realized θ.
2. The central management commits to a set of decision rules, $(\lambda(\cdot), \mu(\cdot), T(\cdot))$.
3. The IS manager sends the message $\hat{\theta} = \sigma_1(\theta)$ to the central management based on the reporting strategy $\sigma_1 : \Theta \rightarrow D_1$.
4. The center makes the decisions $(\lambda(\hat{\theta}), \mu(\hat{\theta}), T(\hat{\theta})) = \sigma_2(\hat{\theta})$ based on the message received and the strategy $\sigma_2 : D_1 \rightarrow D_2$.
5. The IS manager implements the decisions.

A Bayesian equilibrium of this game is a strategy pair (σ_1^*, σ_2^*), the best responses of each to the other's strategy. Because an arbitrary pair of strategies (σ_1, σ_2) induces three outcome functions:

$$\begin{aligned} \lambda : \Theta &\rightarrow \mathcal{R}_+ \\ \mu : \Theta &\rightarrow \mathcal{R}_+ \\ T : \Theta &\rightarrow \mathcal{R}, \end{aligned}$$

these three outcome functions define a mechanism $\Gamma(\cdot) \equiv (\lambda(\cdot), \mu(\cdot), T(\cdot))$. A mechanism therefore is a procedure giving the decisions to the central management, which commits itself to a decision rule relating the choice of Γ to messages sent by the IS manager. That is, a mechanism maps a message space to a set of outcomes, which has the simultaneous purpose of extracting information and making decisions (Guesnerie and Laffont [45]).

A mechanism is *direct* if and only if $D_1 = \Theta$. That is, under a direct mechanism, the IS manager's strategy is a reporting strategy that is a mapping $\sigma_1 : \Theta \rightarrow \Theta$. In order for the IS manager to choose truthful revelation, a mechanism must be *incentive compatible.* Let $U(\hat{\theta};\theta)$ be the IS manager's indirect utility function when the true cost parameter is θ and she reports $\hat{\theta} = \sigma_1(\theta)$. A direct mechanism is *incentive compatible* (a *direct revelation mechanism*) if and only if

$$U(\theta;\theta) \geq U(\hat{\theta};\theta), \qquad \forall \theta, \hat{\theta} \in \Theta.$$

A direct revelation mechanism thus induces the IS manager's reporting strategy to be an identity function $\theta = \sigma_1(\theta)$, for all θ in Θ.

By invoking the well-known *revelation principle* (Dasgupta et al. [21]; Harris and Townsend [46]; Myerson [83]), we know that, for any Bayesian equilibrium (σ_1, σ_2) of any mechanism, there exists a Bayesian equilibrium $(\tilde{\sigma}_1, \sigma_2)$ of a direct mechanism where $\tilde{\sigma}_1(\theta) = \theta$, for all θ in Θ, such that the induced outcomes coincide. That is, given the IS manager's reporting strategy σ_1, the central management can always prescribe another composite mechanism $\tilde{\sigma}_1 = \psi \circ \sigma_1$, mimicking the IS manager's optimization, so that $\tilde{\sigma}_1(\theta) = \psi \circ \sigma_1(\theta) = \theta$. So, any mechanism is isomorphic to a direct mechanism by which the IS manager reveals her information truthfully. Thus, by the revelation principle, we can restrict our attention only to the outcomes induced by a direct incentive compatible mechanism without loss of generality.

I assume that the central management does not have perfect information about $C(\mu, \theta)$, the IS department's cost to operate a system with "effective" capacity μ. Depending on the problem, this cost can have two interpretations: (1) the cost of achieving a particular level of effective capacity for a new system that the organization considers acquiring to replace or to upgrade an existing system in order to accommodate increased demand; or (2) the cost to maintain or improve the effective capacity of the current system. This cost is assumed to depend on the quality of the staff and the ability of the manager of the IS department, which is summarized by the cost parameter θ. For the short-run interpretation, for instance, we may assume that a system with a "raw" (MIPS) capacity $\overline{\mu}$ is already in place and that it will cost the organization $C(\overline{\mu}, \mu, \theta)$ to operate this system at an effective capacity μ, where $C(\overline{\mu}, \mu, \theta)$ is increasing in μ and θ, and $\lim_{\mu \to \overline{\mu}} C(\overline{\mu}, \mu, \theta) \to \infty$.

To limit analytical complexity, I restrict myself to cases where the private information of the IS manager has only one dimension. In $C(\mu, \theta)$, θ is the one-dimensional parameter that captures the central management's uncertainty about both the IS department's productivity (or quality) and the IS manager's ability. Depending on the setting, θ may represent the long-run marginal (effective) capacity cost or the marginal maintenance cost. It is assumed that, for a targeted effective capacity μ, a larger realized θ requires a higher cost to achieve that effective capacity. I impose the following standard assumption on the information structure of the environment:

ASSUMPTION 1 The IS manager knows the realized θ exactly whereas the central management only has some common knowledge prior beliefs about the distribution of $\theta, F(\tilde{\theta})$, which is twice continuously differentiable and has the probability density function $f(\tilde{\theta}) > 0$ if and only if $\tilde{\theta} \in \Theta \equiv [\underline{\theta}, \overline{\theta}]$.[4]

If there exists an external information service market, the organization's IS-related decisions may be affected regardless of which control structure is employed. When the internal IS department's cost is correlated with the market price, the central management may learn more about its IS department's cost from this extra signal. Furthermore, the presence of external competition may in effect reduce the IS manager's informational rent even without actual outsourcing (Caillaud [13]; Lewis and Sappington [64]). To avoid these external market effects, both the users and the IS department are not allowed to access the market. This assumption is reasonable if the organization's infor-

[4]Since the IS personnel's turnover rate is very high and computer technology advances rapidly, the variations of the real θ represented by the central management's prior beliefs reflect the central management's deficient knowledge of computing technologies and the local operating environment.

mation processing requirements are significantly idiosyncratic. Assumption 2 gives the properties of the cost function:

ASSUMPTION 2 For all $\theta \in \Theta$ and $\mu \in \mathcal{R}_+, C(\mu, \theta)$ is assumed to be twice continuously differentiable with respect to both arguments, and $C_\mu > 0, C_\theta > 0, C_{\mu\theta} > 0$, and $C_{\mu\mu} \geq 0$.

The cost function is increasing in both arguments and convex in μ. $C_{\mu\theta} > 0$ merely requires the marginal capacity cost to be monotone in θ, the "sorting condition."

ASSUMPTION 3 The effective capacity can be verified by the central management ex post.

In general, this is a reasonable assumption, since the central management can verify the effective capacity by historical data on average job turnaround time and system availability. The central management can therefore direct the IS manager to achieve a specified level of effective capacity.

Although the IS manager's effort is not explicitly incorporated into the cost function as is usual in principal-agent models, we can transform a cost function with manager's effort explicitly incorporated into our formulation. To see this, let $h(a, \theta)$ be a univariate measure of efficiency units of inputs, which depends on the IS manager's effort, a, and the quality of the IS department, θ. Let $\mu(h(a, \theta))$ be the effective capacity, which is monotone in $h(a, \theta)$. If $\psi(a, \theta)$ is the pecuniary cost to the manager of her effort a, then the cost of the effective capacity μ incurred by the IS department is

$$C(\mu, \theta) = \min_a \ \psi(a, \theta)$$

subject to

$$\mu(h(a, \theta)) = \mu.$$

The manager's effort can therefore be entirely suppressed in the notation. A similar approach is used to deal with stochastic outputs in Laffont and Tirole [61] and McAfee and McMillan [70].

ASSUMPTION 4 The central management and the users have the same gross value function and delay cost, which are the same as in the previous section with no uncertainty involved, and are common knowledge for both the central management and the IS manager.

Since the central management does not know the exact cost parameter θ, it must rely on the IS department's cost report or budget proposal to make corresponding decisions. Thus, when the IS manager's objective does not coincide with that of the central management, the central management faces a constrained optimization problem that maximizes the system's net value and at the same time induces the IS manager's truthful revelation of θ.

In my model, if the IS manager does not derive any utility from consuming "organizational slack" and is paid a flat salary, there is no incentive for the manager to misrepresent her private cost information.[5] But if the IS manager can derive utility from consuming the organizational slack by acquiring superfluous technologies and staff, then the manager has an incentive to exaggerate the IS department's costs. Notice that regardless of whether the IS manager intends to maximize her reward from cost savings or her utility from the consumption of the organizational slack, the optimal strategy for the IS manager is always to exaggerate the costs in order to induce the central management to agree on as large a budget allocation as possible. Without providing incentives to the manager for cost savings, the only instruments that the central management can use to influence the manager's behavior are the centralized decisions on system capacity and budget allocation. Notice that, for the current case, there is no need to specify the IS manager's utility function so long as the IS manager maximizes the excess budget

[5]Organizational slack is "the excess of resources allocated over the minimum necessary to accomplish the tasks assigned" (Antle and Eppen [2]; Cyert and March [20]). In the presence of information asymmetry and objective conflicts between the central management and the departmental or divisional manager, incurring organizational slack is inevitable in general when the divisional manager is given the responsibility to perform certain tasks (see e.g., Antle and Eppen [2]; Antle and Fellingham [3]; Kirby et al. [57]). Organizational slack is here defined as the budget allocation in excess of the minimum necessary to operate the information system at a certain effective capacity, and I assume that the IS manager can derive utility from consuming organizational slack.

allocation.

By taking a mechanism design approach, it is first necessary for the central management to commit credibly to a (direct) revelation mechanism defined by the three outcome functions:

$$(\lambda(\cdot), \mu(\cdot), T(\cdot)).$$

I first derive the required (incentive compatible) amount of excess budget allocation for each $\theta \in \Theta$. I define the excess budget allocation when the IS manager reports $\hat{\theta}$ and the true cost parameter is θ:

$$S(\hat{\theta};\theta) \stackrel{\text{def}}{=} R(\lambda(\hat{\theta}), \mu(\hat{\theta})) - C(\mu(\hat{\theta}), \theta) + T(\hat{\theta}),$$

where $R(\cdot)$ is the IS department's revenue generated from its computing services and the transfer, $T(\cdot)$, can be positive (subsidy) or negative (taxation). To induce the IS manager to report truthfully requires:

$$S(\theta;\theta) \geq S(\hat{\theta};\theta), \qquad \forall \hat{\theta}, \theta \in \Theta, \tag{2.7}$$

i.e., revealing the true cost information is her (weakly) dominant strategy.

Of course, if the transfer rule (2.6) can automatically induce the IS manager to report truthfully, then the full-information solution can always be obtained and the private information of the IS manager imposes no informational constraint on the central management. However, when the IS manager's objective is to maximize the excess budget allocation, the transfer rule (2.6) fails to be incentive compatible, as shown by the following proposition:

PROPOSITION 2.2 *If $C(\mu, \theta)$ is increasing in θ for all $\theta \in \Theta$, then the subsidy rule (2.6) is not incentive compatible, and it is therefore impossible to achieve the full-information solution since the IS manager will exaggerate the IS department's costs.*

PROOF. I prove this proposition in its most general form with no differentiability assumed and show that (2.6) does not satisfy (2.7). Consider any two distinct possible realizations of θ: $\theta_1, \theta_2 \in \Theta$ with $\theta_2 > \theta_1$. Let $\lambda_i \equiv \lambda(\theta_i)$, $\mu_i \equiv \mu(\theta_i)$, $R_i \equiv R(\lambda_i, \mu_i)$, and $T_i \equiv T(\theta_i)$ for $i = 1,2$. Inducing the IS manager's truth-revelation requires

$$\begin{aligned} S(\theta_1;\theta_1) &\geq S(\theta_2;\theta_1) \\ S(\theta_2;\theta_2) &\geq S(\theta_1;\theta_2) \end{aligned}$$

which, by the definition of $S(\cdot)$, together with the subsidy rule (2.6), which says $T_i = C(\mu_i,\theta_i) - R_i$ for $i = 1,2$, imply

$$\begin{aligned} 0 &\geq C(\mu_2,\theta_2) - C(\mu_2,\theta_1) \\ 0 &\geq C(\mu_1,\theta_1) - C(\mu_1,\theta_2) \end{aligned}$$

But $C(\mu,\theta)$ is increasing in θ, and since $\theta_2 > \theta_1$, $C(\mu_2,\theta_2) > C(\mu_2,\theta_1)$ and $C(\mu_1,\theta_2) > C(\mu_1,\theta_1)$. Hence, the first inequality is violated, and thereby if the central management follows the rule (2.6), the manager will exaggerate the realized θ. ||

Although the IS manager receives utility from consuming $S(\theta)$, she will not in general find it to be a perfect substitute for a cash payment, since there will be some restrictions on how $S(\theta)$ can be spent. As a result, the organization may be better off if it can induce truthful reporting by paying cash for cost savings rather than having the IS manager consume all of $S(\theta)$. With general functional forms for the IS manager's utility generated from consuming organizational slack and pecuniary rewards, however, the problem becomes intractable. In order to sharpen the results and obtain a fuller characterization of the incentive compatible mechanisms, I assume that the IS manager has a utility function linear in both pecuniary rewards and consumption of organizational slack. Henceforth I make the following assumption about the IS manager's utility function.

ASSUMPTION 5 The IS manager has a common knowledge utility function:

$$\tilde{U}(Z,\hat{\theta}) = \max_{0\leq\phi\leq Z} \{B(\phi,\hat{\theta}) + \xi\,(Z-\phi)\},$$

where

$B(\phi,\hat{\theta})$: the pecuniary reward to the IS manager if she shows a cost savings (a profit) ϕ and reports $\hat{\theta}$ when the IS department is organized as a cost center (a profit center);

ϕ : the amount of excess budget allocation (profit) that the IS manager chooses to be shown as a cost savings (profit);

ξ : the index of the strength of the IS manager's professionalism tendency, $\xi \in [0,1]$.

I restrict ξ within $[0,1]$, since for $\xi > 1$, it is never worth the central management's while to lure the IS manager away from consuming organizational slack. Note that given this assumption, if the IS manager is not rewarded for a cost savings, then $B(\phi, \hat{\theta}) = 0$ for all ϕ in $[0, Z]$, and if consuming organizational slack does not generate any positive utility for the IS manager, $\xi = 0$.[6]

Given $B(\cdot), Z$, and ξ, the IS manager derives utility $B(\phi(Z), \hat{\theta})$ from pecuniary rewards, and utility $\xi(Z - \phi(Z))$ from excess investment, respectively, when she chooses to show cost savings (profit) by an amount of $\phi(Z)$. This linear structure of the IS manager's utility function is restrictive, but for expositional simplicity it can be viewed as a first-order approximation of a more general utility function. Besides, my main aim here is to demonstrate the IS manager's substitution between consuming organizational slack and pecuniary rewards. I believe that most of the qualitative results derived from the simple model still hold with a more general utility function. As long as the IS manager's tendency to professionalism is weak enough that a mixture of consuming slack and receiving pecuniary rewards gives a higher utility than consuming the organizational slack alone, the central management should provide some pecuniary rewards for a cost savings to encourage the IS manager's substitution.

For the moment, I assume that ξ is known exactly by the central management. In Section 3.7 I relax this assumption by assuming that the central management only has some prior beliefs about the actual ξ. Whether or not ξ is known to the central management, my mechanisms do not require the IS manager to reveal her actual ξ for two reasons. First, mechanism design problems with more than one informational parameter

[6] $1 - \xi$ may be considered as the "welfare" weights that the central management puts on the IS manager's overconsumption. This interpretation is usual in studies dealing with regulating a monopolist; see, e.g., Baron and Myerson [9]. Alternatively, the IS manager may able to "siphon off" a fraction ξ of the excess budget allocation or profit, as in Hart [48].

are extremely difficult to solve even with very simple utility functions. The difficulty arises because the set of partial differential equations characterizing the incentive compatible mechanisms are very hard to characterize (Baron [7]). Second, since the job of central management is to be expert in managing *people*, over time they should be able to form a reasonable opinion as to ξ. The central management cannot reasonably be expected to be expert with regard to running an information system, so it makes sense to focus on θ rather than ξ when only one informational parameter is involved.

Notice that because the central management values the extra allocation $S(\hat{\theta};\theta)$ fully but the IS manager either consumes the extra allocation (i.e., $\phi = 0$) to get $\xi S(\hat{\theta};\theta)$ or shows a cost savings $S(\hat{\theta};\theta)$ (i.e., $\phi = S$) to get a pecuniary reward $B(S(\hat{\theta};\theta),\hat{\theta})$ or both (i.e., $0 < \phi(S) < S$), then, given $B(\cdot)$ and $S(\hat{\theta};\theta)$, the loss to the organization due to the extra allocation is equal to:

$$B(\phi(S(\hat{\theta};\theta),\hat{\theta}) + S(\hat{\theta};\theta) - \phi(S(\hat{\theta};\theta)).$$

I assume that, when the IS manager is indifferent toward the pecuniary reward and the consumption of organizational slack, she chooses the pecuniary reward. Given the linear structure of the IS manager's utility function, the optimal pecuniary reward function $B(\cdot)$ is linear in the IS department's cost savings.

LEMMA 2.1 *Given the excess budget allocation S, it is always optimal for the central management to induce the IS manager to choose $\phi^*(S) = S$, and the optimal rewards on cost savings $B^*(S,\hat{\theta}) = \xi S$.*

PROOF. From the linear structure of the IS manager's utility function, it is obvious that $\phi(S) = S$ if $B(S,\hat{\theta}) \geq \xi S$ and that $\phi(S) = 0$ if $B(S,\hat{\theta}) < \xi S$. Then because the central management values $S(\theta)$ fully and by assumption $\xi \in [0,1]$, it is obvious that $B(\phi) = \xi \phi$, and the central management always induces the IS manager to take the reward from cost savings (i.e., $\phi^*(S) = S$). ||

Given the IS manager's incentive problems, the central management will seek to design a set of performance criteria and decision rules that restructures the IS manager's objective function and thereby reduces or eliminates the informational rent that the IS manager can command. Although I focus on cases with a single service center, my result can easily be extended to the open network case (Wang and Sumita [109]).

Chapter 3

Unlimited Communication and Feasible Systems

3.1 Introduction

Although it has been shown that it is never optimal for organizations to organize an IS department as a profit center due to the monopolistic pricing problem, this conclusion is derived from models having neither information asymmetry nor objective conflicts between the central management and the IS department (Dewan and Mendelson [25]; Mendelson [72]). As it turns out, when communication is unlimited, the same conclusion results in a simple way from my model as well, since the revelation principle implies that a suitably designed revelation mechanism can replicate the performance of any decentralized mechanism such as a profit center, and an optimally designed revelation mechanism is dominated by no other method of control. I first focus on deriving the optimal revelation mechanism. I assume in this chapter that the set of feasible systems is continuous, i.e., $\mathcal{K} = \mathcal{R}_{+}$.

3.2 Revelation Mechanisms without Rewards for Cost Savings

Given the mechanism committed to by the central management and the realized cost parameter θ, the IS manager maximizes the excess budget allocation by solving the following problem:

$$\max_{\hat{\theta}} S(\hat{\theta};\theta),$$

which gives the first-order condition:

$$S_{\hat{\theta}}(\hat{\theta};\theta) = 0, \tag{3.1}$$

and the second-order condition:

$$S_{\hat{\theta}\hat{\theta}}(\hat{\theta};\theta) \leq 0. \tag{3.2}$$

Since the IS manager already knows the realized θ, the revenue plus the transfer must at least fully cover the costs in order for the IS manager to agree to the terms of the budget allocation; i.e.,

$$S(\theta;\theta) \geq 0, \qquad \forall\theta \in \Theta. \tag{3.3}$$

I call this the budget constraint, which is similar to the individual rationality constraint in the incentive literature. However, here this budget constraint simply serves as a yardstick for the central management when calculating the appropriate taxation from or subsidy to the IS department so that the IS department can at least balance its budget for all $\theta \in \Theta$. Thus (2.7) and (3.3) form a set of constraints for the optimal mechanism design problem, and the set of mechanisms satisfying (2.7) and (3.3) is feasible. Lemma 3.1 now can characterize the set of feasible mechanisms. This lemma is well-known in incentive literature (see, e.g., Baron and Myerson [9]; Guesnerie and Laffont [45]; Mirrlees [78]).

LEMMA 3.1 *A direct, differentiable mechanism* (λ, μ, T) *is incentive compatible if and only if*

$$S(\theta) = S(\overline{\theta}) + \int_{\theta}^{\overline{\theta}} C_{\theta}(\mu(\tilde{\theta}), \tilde{\theta})d\tilde{\theta}, S(\overline{\theta}) \geq 0, \tag{3.4}$$

and $\mu(\theta)$ *is non-increasing for all* θ *in* Θ.

PROOF. Local incentive compatibility requires

$$S(\theta;\theta) \geq S(\hat{\theta};\theta), \qquad \forall\hat{\theta},\theta \in \Theta,$$

and (by reversing the role of θ and $\hat{\theta}$)

$$S(\hat{\theta};\hat{\theta}) \geq S(\theta;\hat{\theta}), \qquad \forall\hat{\theta},\theta \in \Theta,$$

which imply

$$-C(\mu(\theta),\theta) + C(\mu(\theta),\hat{\theta}) \geq S(\theta;\theta) - S(\hat{\theta};\hat{\theta}) \geq -C(\mu(\hat{\theta}),\theta) + C(\mu(\hat{\theta}),\hat{\theta}).$$

For $\theta > \hat{\theta}$, dividing through the inequalities by $\theta - \hat{\theta}$ and taking the limit as $\hat{\theta} \to \theta$ give:

$$S'(\theta) = -C_\theta(\mu(\theta),\theta).$$

Then (3.4) is obtained by integration.

Since the first-order condition, $S_{\hat{\theta}}(\theta;\theta) = 0$, holds as an identity, total differentiation yields:

$$S_{\hat{\theta}\hat{\theta}}(\theta;\theta) + S_{\hat{\theta}\theta}(\theta;\theta) = 0.$$

Thus, the local second-order condition is equivalent to requiring

$$S_{\hat{\theta}\theta}(\theta;\theta) = -C_{\theta\mu}(\mu(\theta),\theta)\frac{d\mu(\theta)}{d\theta} \geq 0.$$

But by assumption $C_{\theta\mu} > 0$, and thereby it is necessary that $\frac{d\mu(\theta)}{d\theta} \leq 0$.

To prove the sufficient part of the lemma, suppose that

$$S(\theta_2;\theta_1) > S(\theta_1;\theta_1),$$

for some θ_1 and θ_2. Then this implies

$$\int_{\theta_1}^{\theta_2} S_{\hat{\theta}}(\tilde{\theta};\theta_1)d\tilde{\theta} > 0.$$

If $\theta_2 > \theta_1$,

$$S_{\hat{\theta}}(\tilde{\theta};\theta_1) \leq S_{\hat{\theta}}(\tilde{\theta};\tilde{\theta}) = 0$$

for $\tilde{\theta} \geq \theta_1$ since $S_{\hat{\theta}\theta}(\hat{\theta};\theta) \geq 0$. We obtain a contradiction. The case with $\theta_1 > \theta_2$ can be proved similarly. ||

Since the central management can do no better than allocating an amount of excess budget equal to $S(\theta)$ to induce the IS manager's truthful revelation, $S(\theta)$ may be viewed as the IS manager's informational rent. That is, the central management must allocate some organizational slack in order to induce the IS manager's truth-revelation if consuming organizational slack can generate positive utility for the IS manager. Furthermore, the monotonicity constraint that $\mu(\theta)$ must be non-increasing can be explained intuitively. Note that by assumption $C(\mu, \theta)$ is increasing in θ. Thus, if $\mu(\theta)$ is increasing in θ, the IS manager's dominant strategy is to report $\hat{\theta} = \overline{\theta}$; i.e., she will report her department's costs as high as possible. Consequently, it is impossible for the central management to construct a mechanism that can induce the IS manager's truth-revelation.

From a technical point of view, the purpose of this lemma is twofold: first, it gives a property of the excess budget allocation implied by (2.7); second, this property can be used to replace the constraints (3.3) by a single constraint (Baron [7]). Notice that $S(\theta)$ is decreasing in θ since $C_\theta > 0$. Since the excess budget allocation is undesirable to the central management, it should set $S(\overline{\theta}) = 0$. By setting $S(\overline{\theta}) = 0$, (3.3) is automatically satisfied for every θ in Θ without disrupting the incentive compatibility constraints. Intuitively the IS department likes to overstate its costs and the incentive constraint is therefore upward-binding. To prevent the IS department from overstating its cost for all θ in Θ, $S(\theta)$ must be at least as large as $S(\tilde{\theta}; \theta)$ for all $\tilde{\theta} \in (\theta, \overline{\theta}]$. When the state is $\overline{\theta}$, there is no other state that the IS department can mimic, so the informational rent equals zero; i.e., $S(\overline{\theta}) = 0$.

From Lemma 3.1, given a non-increasing capacity setting rule $\mu(\cdot)$, as long as the excess budget allocated to the IS department equals $S(\hat{\theta})$ when the manager reports $\hat{\theta}$, it is always optimal for the manager to report truthfully. The central management's problem then reduces to finding the expected net value maximizing capacity function $\mu(\cdot)$ within the set of non-increasing functions.

Without rewarding the IS manager for cost savings, the central management will incur an extra cost equal to $S(\theta)$. However, if the IS manager values pecuniary rewards, it may be advantageous for the central management to compensate the IS manager for cost savings to offset the IS manager's incentive to consume all or part of the extra budget allocated. Since cases without an incentive scheme for cost savings can be considered as a special case of cases with an incentive scheme, I do not characterize the optimal mechanism here. The optimal mechanism for cases with an incentive scheme are the subject of the next section.

3.3 Revelation Mechanisms with Rewards for Cost Savings

Define the IS manager's indirect utility function

$$\begin{aligned} U(\hat{\theta};\theta) &\equiv \tilde{U}(S(\hat{\theta};\theta),\hat{\theta}) \\ U(\theta) &\equiv U(\theta;\theta). \end{aligned}$$

To clarify the presentation, I suppress ξ from the IS manager's utility function. From Lemma 2.1, the optimal pecuniary reward to the IS manager for cost savings then equals a fraction, ξ, of the cost savings that the IS department shows, and $\phi^*(S(\theta)) = S(\theta)$ for all θ in Θ. Thus the IS manager's indirect utility function can be written as:

$$U(\hat{\theta};\theta) = \xi S(\hat{\theta};\theta).$$

Since the central management seeks to maximize the expected organizational net value generated by the information services, based on the queueing model, the mechanism design problem is:

$$\max_{\lambda(\cdot),\mu(\cdot),S(\cdot)} \int_{\Theta} \{V(\lambda(\theta)) - \lambda(\theta)W(\lambda(\theta),\mu(\theta)) - C(\mu(\theta),\theta) - \xi S(\theta)\}\, dF(\theta) \tag{3.5}$$

subject to

$$U(\theta) \geq U(\hat{\theta};\theta), \qquad \forall \theta, \hat{\theta} \in \Theta \tag{3.6}$$

$$U(\theta) \geq 0, \qquad \forall \theta \in \Theta. \tag{3.7}$$

By substituting $\xi S(\theta)$ for $S(\theta)$ in Lemma 3.1, (3.6) and (3.7) are satisfied if $U(\overline{\theta}) \geq 0$ and

$$U(\theta) = U(\overline{\theta}) + \xi \int_{\theta}^{\overline{\theta}} C_{\theta}(\mu(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta},$$

provided $\mu(\theta)$ is non-increasing in θ. When $\xi > 1$, $\xi S > S$, so the central management will never provide the IS manager pecuniary rewards for a cost savings if $\xi > 1$. Consequently, the case without an incentive scheme mentioned above corresponds to cases where $\xi \geq 1$. Since it is always optimal from the central management's standpoint to set $U(\overline{\theta}) = 0$, the organizational loss is:

$$U(\theta) = \xi S(\theta) = \xi \int_{\theta}^{\overline{\theta}} C_\theta(\mu(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta}. \tag{3.8}$$

Using (3.8) in (3.5) and integrating by parts, the mechanism design problem can be recast as:

$$\max_{\lambda(\cdot),\mu(\cdot)} \int_{\Theta} \{V(\lambda(\theta)) - \lambda(\theta)W(\lambda(\theta), \mu(\theta)) - C(\mu(\theta), \theta) - \xi\beta(\theta)C_\theta(\mu(\theta), \theta)\} dF(\theta) \tag{3.9}$$

subject to

$$\mu(\theta) \text{ is non-increasing in } \theta, \ \forall \theta \in \Theta, \tag{3.10}$$

where $\beta(\theta) \equiv \frac{F(\theta)}{f(\theta)}$ is the (inverse) hazard rate. The usual solution approach is to solve the unconstrained problem (3.9) first, and then to check whether or not (3.10) is satisfied for all θ. If it is not satisfied for all θ (i.e., when bunching or pooling occurs), then some convexification techniques developed by Baron and Myerson [9] or Guesnerie and Laffont [45] must be used. However, adopting these techniques to solve the problem when constraint (3.10) is binding greatly complicates the mathematics while providing little insight into this problem. In the following proposition, therefore, I impose some assumptions on (3.9) so that (3.10) can be satisfied for every θ:

Proposition 3.1 *If*

1. *The Hessian matrix of*

$$H \stackrel{\text{def}}{=} V(\lambda) - \lambda W(\lambda, \mu) - C(\mu, \theta) - \xi\beta(\theta)C_\theta(\mu, \theta)$$

with respect to λ and μ is negative definite;

2. *$C_\mu(\mu, \theta) + \xi\beta(\theta)C_{\theta\mu}(\mu, \theta)$ is non-decreasing in θ for all $\theta \in \Theta$,*

the optimal solution of (3.9)–(3.10) is the solution of the following equations:

$$0 = V'(\lambda) - W(\lambda,\mu) - \lambda W_\lambda(\lambda,\mu) \tag{3.11}$$

$$0 = -\lambda W_\mu(\lambda,\mu) - C_\mu(\mu,\theta) - \xi\beta(\theta)C_{\theta\mu}(\mu,\theta) \tag{3.12}$$

and the optimal solution is globally incentive compatible. Moreover, letting $\lambda^*(\theta)$ and $\mu^*(\theta)$ be the solution of (3.11) and (3.12), the optimal price $p^*(\theta) = \lambda^*(\theta)W_\lambda(\lambda^*(\theta),\mu^*(\theta))$, and the optimal incentive compatible transfer function is:

$$T^*(\theta) = -p^*(\theta)\lambda^*(\theta) + C(\mu^*(\theta),\theta) + S^*(\theta),$$

where

$$S^*(\theta) = \int_\theta^{\bar\theta} C_\theta(\mu^*(\tilde\theta),\tilde\theta)d\tilde\theta.$$

PROOF. Maximizing (3.9) pointwise with respect to λ and μ gives the following first-order conditions:

$$\frac{\partial H}{\partial \lambda} = V'(\lambda) - W(\lambda,\mu) - \lambda W_\lambda(\lambda,\mu) = 0$$

$$\frac{\partial H}{\partial \mu} = -\lambda W_\mu(\lambda,\mu) - C_\mu(\mu,\theta) - \xi\beta(\theta)C_{\theta\mu}(\mu,\theta) = 0$$

Letting $\mu^*(\theta)$ denote the optimal capacity, I now show that $\frac{d\mu^*(\theta)}{d\theta} \leq 0$ provided the assumptions hold. Let $\mathcal{H}$ be the Hessian matrix of H, and by the implicit function theorem (see, e.g., Silberberg [99] p. 144),

$$\left[\frac{d\lambda^*(\theta)}{d\theta} \quad \frac{d\mu^*(\theta)}{d\theta}\right]^T = -\mathcal{H}^{-1}[H_{\lambda\theta} \quad H_{\mu\theta}]^T,$$

where the superscript T denotes the transpose operation. Writing out the right-hand side of this equation gives:

$$-\frac{1}{|\mathcal{H}|}\begin{bmatrix} H_{\mu\mu} & -H_{\mu\lambda} \\ -H_{\lambda\mu} & H_{\lambda\lambda} \end{bmatrix}\begin{bmatrix} H_{\lambda\theta} \\ H_{\mu\theta} \end{bmatrix},$$

where $|\mathcal{H}|$ is the determinant of $\mathcal{H}$. Since $\mathcal{H}$ is assumed to be negative definite, $|\mathcal{H}|$ is positive. Because only $\mu^*(\theta)$ is required to be non-increasing, the incentive constraint is satisfied if $-H_{\lambda\mu}H_{\lambda\theta} + H_{\lambda\lambda}H_{\mu\theta} \geq 0$. But $H_{\lambda\theta} = 0$ and $H_{\lambda\lambda} < 0$, and thereby $\mu(\theta)$ is non-increasing if $H_{\mu\theta} \leq 0$. But, $H_{\mu\theta} = -C_{\theta\mu}(1 + \xi\beta'(\theta)) - \xi\beta(\theta)C_{\theta\mu\theta}$, and thereby (3.10) is satisfied provided that the second assumption holds.

Finally, from Lemma 3.1, the optimal mechanism is globally incentive compatible if $\mu^*(\theta)$ is non-increasing in θ. The optimal price follows directly from (3.11). ||

Since $H_{\lambda\theta} = 0$, the proof of Proposition 3.1 implies that the signs of both $\frac{d\lambda^*(\theta)}{d\theta}$ and $\frac{d\mu^*(\theta)}{d\theta}$ are determined by the sign of $H_{\mu\theta}$. Because $H_\mu = C_\mu(\mu,\theta) + \xi\beta(\theta)C_{\theta\mu}(\mu,\theta)$, the *virtual* marginal capacity cost, both $\lambda^*(\theta)$ and $\mu^*(\theta)$ are decreasing if the virtual marginal capacity cost is increasing in θ. Given $C_{\theta\mu\theta} \geq 0$ and the assumption that $C_{\mu\theta} > 0$, the only complication that can arise to make the sign of $H_{\mu\theta}$ ambiguous is the sign of $\beta'(\theta) = \frac{f(\theta)^2 - F(\theta)f'(\theta)}{f(\theta)^2}$. So $\beta(\theta)$ is increasing if and only if $f(\theta)^2 - F(\theta)f'(\theta) > 0$, i.e., if and only if, for all θ:

$$\frac{F(\theta)}{f(\theta)} < \frac{f(\theta)}{f'(\theta)}.$$

This condition, the monotone hazard rate condition, is implied by the well-known monotone likelihood ratio property. Thus, whenever the distribution function satisfies the monotone likelihood ratio property, the inverse hazard rate $\beta(\theta)$ is increasing, and thereby the monotonicity constraint (3.10) is satisfied for all θ. This property is satisfied by many families of probability distributions, such as Normal, Exponential, Poisson, and Uniform (Milgrom [77]).

Discussion. The way the optimal mechanism works can be visualized as follows: the central management first displays a menu of values determined from the mechanism, $\{\Gamma^*(\theta)) : \theta \in \Theta\}$, to the IS manager; the central management then asks the IS manager to pick a particular entry from the menu; after the IS manager picks the entry, the decisions and actions are implemented accordingly without any ex post adjustments. The credibility of the central management's commitment is crucial here. Within an organization, the central management's ability to make a credible commitment can be supported by an established performance evaluation system and the importance of maintaining a reputation of good faith insofar as future negotiations with the IS manager as well as all other managers are concerned. If the central management fails to convince the IS manager that it will not use the additional information revealed by her choice against her ex post, the IS manager will not reveal her private information truthfully, so that each party will be facing a game similar to the original one and the mechanism must be posterior implementable (Green and Laffont [41]).

If the central management can credibly commit itself to a mechanism characterized by

$$(\lambda^*(\theta), \mu^*(\theta), T^*(\theta)),$$

it is optimal for the IS manager to report θ truthfully. Since only $\mu^*(\theta)$ is involved in $S^*(\theta)$, it is the central management's choice of capacity (and thereby the corresponding budget allocation) that dictates the IS manager's reporting strategy. Without loss of generality, the mechanism can be considered to tax away all the IS department's revenue and then allocate it a lump-sum budget equal to:

$$C(\mu^*(\theta), \theta) + S^*(\theta).$$

With this approach, the IS manager's informational rent is:

$$U(\hat{\theta};\theta) = \xi \left\{ C(\mu^*(\hat{\theta}), \hat{\theta}) + \int_{\hat{\theta}}^{\bar{\theta}} C_\theta(\mu^*(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta} - C(\mu^*(\hat{\theta}), \theta) \right\}$$

if the IS manager reports $\hat{\theta}$. The incentive compatibility of the optimal budget allocation rule can then be easily checked by observing that

$$\left. \frac{\partial U(\hat{\theta};\theta)}{\partial \hat{\theta}} \right|_{\hat{\theta}=\theta} \equiv 0.$$

So as long as the central management and the IS manager have the same amount of information concerning the users' demand, the IS manager's incentive for truth-revelation will not be altered even when the realized revenue differs from $p^*(\theta)\lambda^*(\theta)$. Also note that the outcomes of this mechanism are equivalent to the outcomes of a bargaining game when the central management has all the bargaining power, and the outcomes of the optimal mechanism correspond to the central management's *neutral bargaining solution* (see Myerson [84]). Thus, even when the central management has superior information about the users' demand or its own desired scale of IS operations, the optimal mechanism remains optimal with some appropriate modifications in response to a different value function.

The sources of the distortions from both the incentive conflicts and the information asymmetry are clearly visible in equations (3.11) and (3.12). First, comparing (3.11) and its full-information counterpart, equation (2.2), it is clear that there is no direct distortion in the short-run problem. That is, given a particular capacity, both (2.2) and (3.11) will yield the same solution. However, comparing (2.3) and (3.12), it is clear that the full-information first-order condition is distorted by the extra term $\xi\beta(\theta)C_{\theta\mu}(\mu,\theta) \geq 0$. This has the effect of reducing the optimal capacity from that of the full-information solution, which in turn will reduce the arrival rate via (3.11).

The intuition behind the restriction in capacity can be seen from the fact that the IS manager's ex post informational rent is $\xi S(\theta) = \xi \int_{\theta}^{\overline{\theta}} C_{\theta}(\mu(\tilde{\theta}),\tilde{\theta})d\tilde{\theta}$. Since $C_{\theta\mu} > 0$, distorting μ downward will reduce the integrand in $\xi S(\theta)$, reducing her informational rent and her incentive to overstate cost. The ex post informational rent is calculated based upon the mechanism (note $\mu(\theta)$ in the integrand), and, as can be seen from the limits of integration, it is zero at $\overline{\theta}$ and at its maximum at $\underline{\theta}$. This of course is to be expected, since it is impossible for the IS manager to misrepresent the least efficient department (i.e., $\overline{\theta}$) as anything else, and thus she has no power to extract any reward for her information. On the other hand, she has the greatest latitude for misrepresentation when the department is most efficient (i.e., $\underline{\theta}$), and thus must be rewarded most to induce truthful reporting.

It is also interesting to compare the ex post informational rent to the virtual informational rent, the term $\xi\beta(\theta)C_{\theta}(\mu,\theta)$ in H. Intuitively, the role of the virtual informational rent is to cause the mechanism to be designed to guard against reports of large values of θ. When $\theta = \underline{\theta}$, there is no other state in Θ that the IS manager can overstate it to be since $F(\underline{\theta}) = 0$. Consequently, from the central management's standpoint, distorting $\mu(\underline{\theta})$ ex ante will affect the IS manager's incentive for misrepresentation with probability zero, and therefore the virtual informational rent should cause no penalty to be placed on such a report; this is called "no distortion at the bottom." However, when $\theta > \underline{\theta}$, a distortion of $\mu(\theta)$ will reduce the IS manager's incentive for misrepresentation for all $\tilde{\theta} \in [\underline{\theta},\theta]$, and distorting $\mu(\overline{\theta})$ can reduce the IS manager's informational rent for all possible realizations of θ. As a result, the central management wants the mechanism to place the highest penalty on the IS manager's reporting $\overline{\theta}$. As was shown earlier, if $\beta'(\theta) > 0$ everywhere on Θ, $\beta(\overline{\theta}) = \frac{1}{f(\overline{\theta})}$ is its maximum, and $\frac{\xi C_\theta(\mu(\overline{\theta}),\overline{\theta})}{f(\overline{\theta})} > 0$.

From the preceding discussion we see that there are two sources of efficiency loss from the organization's standpoint. The first is the compensation $\xi S(\theta)$ to the IS manager, and the second is the opportunity cost of the jobs not served due to the reduction in arrival rate which is driven by the restriction on capacity. Hence capacity must be set to balance the reduction in informational rent against the loss due to fewer jobs served; this is accomplished through the term $\xi\beta(\theta)C_{\theta\mu}(\mu,\theta)$ in (3.12), the contribution of the virtual informational rent to the virtual marginal capacity cost.

The effects of the incentive conflicts are captured by ξ in (3.12). Note that $\xi = 0$ can be interpreted as meaning that the organization's objectives and those of the IS manager coincide, so that (3.12) reduces to the full information first-order condition, (2.3). As ξ grows, so does the impact of $\xi\beta(\theta)C_{\theta\mu}(\mu,\theta)$, up to the point $\xi = 1$, and the IS manager consumes all of the slack for $\xi \geq 1$. Thus as ξ grows, the central management will find it optimal to distort the capacity downward progressively more, and the optimal arrival rate with it.

The degree of the central management's uncertainty about the IS department's efficiency is similarly captured by $\beta(\theta) = \frac{F(\theta)}{f(\theta)}$, albeit in a somewhat less obvious way. This is most easily seen by using a uniform distribution defined on the interval $[\underline{\theta}, \overline{\theta}]$, which gives $\beta(\theta) = \theta - \underline{\theta}$, where $\underline{\theta} \leq \theta \leq \overline{\theta}$. The central management's degree of uncertainty is then reflected by the length of $[\underline{\theta}, \overline{\theta}]$. Clearly, when the central management is perfectly informed about θ, $S(\theta) = 0$, so that (3.12) coincides with (2.3), as we would expect.

The expression $\beta(\theta) = \theta - \underline{\theta}$ is interesting because it so clearly emphasizes the rather paradoxical nature of the central management's problem stemming from its uncertainty about θ. When there is a possibility that the IS operation is highly efficient (i.e., $\underline{\theta}$ is very low), this enhances the ability of the IS manager to generate large amounts of informational rent when reporting a given θ, and therefore (3.12) says that the lower $\underline{\theta}$ is, the larger is the required downward distortion of the capacity, since $\beta(\theta)$ is larger.

3.4 Profit Center

As shown below, the revelation principle implies that the performance of a cost center governed by the optimal mechanism is at least as good as a profit center given the nature of the information asymmetry in the model. However, this result is based on the assumption that communication between the central management and the IS manager is unlimited and costless. When communication is more problematic, the central management may be attracted to the profit center as an alternative organizational form. The difference in expected net values between the cost center and profit center can therefore be interpreted as an upper bound on the value of improving communications between IS management and central management. Hence I study the profit center in this section.

As in the cost center case, if the IS department does not obtain any rewards from its profit, it is optimal for the IS manager to consume all the profit and show a zero profit. That is, the organizational slack still can appear in the form of profit. It is therefore again necessary for the central management to provide some incentives to keep the IS manager from consuming all the profit. The central management's decision on how to reward the IS manager will not affect her strategies for determining capacity and prices; she will set the capacity and prices to maximize her department's profit. However, the IS manager's decision on consuming organizational slack (the profit) will be altered. That is, the IS manager will set the capacity and price to maximize her department's profit,

$$\pi(\theta) \equiv \pi(\lambda^p(\theta), \mu^p(\theta)) = \max_{\lambda,\mu} p\lambda - C(\mu, \theta), \tag{3.13}$$

where $p = V'(\lambda) - W(\lambda, \mu)$.

Although the delay costs are not directly borne by the IS department, it must take full account of the users' delay costs when determining its profit maximizing price, since (3.13) is equivalent to:

$$\max_{\lambda,\mu} \lambda V'(\lambda) - \lambda W(\lambda, \mu) - C(\mu, \theta).$$

For a given λ, the IS-related costs borne by the organization when the IS department is organized as a profit center are exactly the same as when the central management is fully informed. However, the IS department's profit maximizing price does not maximize the organizational net value. In contrast with the revelation mechanism, the distortion of decisions here stems from the distortion of the users' gross value, not the cost, since the users' gross value should be evaluated as $V(\lambda)$ instead of $\lambda V'(\lambda)$ when maximizing the organizational net value. Also by assumption $V(\lambda)$ is concave, and thereby $V(\lambda) > \lambda V'(\lambda)$ for all $\lambda > 0$. Thus, as long as the IS department is able to earn a positive profit, the resulting organizational net value will be positive. Because $V(\lambda) = \lambda V'(\lambda)$ for all λ if and only if $V(\lambda)$ is linear, the net value maximizing decisions coincide with the IS department's profit maximizing decisions if and only if $V(\lambda)$ is linear. Thus, for each θ, the solution to (3.13) is not equal to the full-information solution unless $V(\lambda)$ is linear.

By the revelation principle, the expected organizational net value when the IS department is organized as a profit center cannot be greater than that attained by a cost center governed by the optimal revelation mechanism, because an appropriately designed centralized mechanism can replicate the outcomes attained by any decentralized mechanism when *the communication is unlimited and costless.* To replicate the performance of the profit center by a centralized mechanism, the central management only needs to commit itself to the mechanism $\{\lambda^p(\theta), \mu^p(\theta) : \theta \in \Theta\}$ and compensate the IS manager by an amount $\xi\pi(\theta)$. It is clear that this mechanism is incentive compatible, since the decisions made by the central management are the same as those the IS manager would make if the decisions were left to her.

As mentioned above, in order to induce the IS manager not to consume the organizational slack, the central management still must provide appropriate rewards for profit. Given the linearity of the IS manager's utility structure, the optimal rewards take the same form as for the cost center, $\xi\pi(\theta)$. Thus, when the IS department is organized as a profit center, the expected organizational net value is:

$$\begin{aligned} E\{NV^p(\theta)\} &\stackrel{\text{def}}{=} \int_\Theta \{V(\lambda^p(\theta)) - \lambda^p(\theta)W(\lambda^p(\theta), \mu^p(\theta)) - C(\mu^p(\theta), \theta) - \xi\pi(\theta)\}\, dF(\theta) \\ &= \int_\Theta \{V(\lambda^p(\theta)) - \lambda^p(\theta)V'(\lambda^p(\theta)) + (1-\xi)\pi(\theta)\}\, dF(\theta). \end{aligned}$$

Providing appropriate rewards for profit gives the organization a strictly positive expected net gain

$$(1-\xi)\int_\Theta \pi(\theta) dF(\theta),$$

provided that $\xi < 1$. When $\xi \geq 1$, the expected organizational net value is simply the expected difference between the users' and the IS department's valuations.

3.5 Some Comparative Statics

In this section, I perform some comparative static analysis by varying the support of the central management's prior beliefs about the IS department's cost parameter, Θ, and the index of the IS manager's preference concerning the excess budget allocation, ξ. I also parameterize the users' aggregate value function $V(\lambda, k)$ by a parameter k such that $V_k(\lambda, k) > 0$ and study the effect of varying k. Let $H^*(\theta)$ be $H(\theta)$ evaluated at $\lambda = \lambda^*(\theta)$ and $\mu = \mu^*(\theta)$, and therefore $E\{NV^*(\theta)\} = E\{H^*(\theta)\}$. I further assume that $F(\theta)$ is uniform and examine the effects of mean-preserving spreads using the parameter δ, i.e., $\Theta = [\underline{\theta} - \delta, \overline{\theta} + \delta]$.

The Effects of Varying Θ

The Effects of Varying $\underline{\theta}$. Differentiating $E\{H^*(\theta)\}$ with respect to $\underline{\theta}$ gives:

$$\frac{d}{d\underline{\theta}}E\{NV^*(\theta)\} = -H^*(\underline{\theta})f(\theta) + \int_{\underline{\theta}}^{\overline{\theta}} \left\{\frac{dH^*(\theta)}{d\underline{\theta}}f(\theta) + H^*(\theta)f(\theta)^2\right\} d\theta \quad (3.14)$$

$$= (E\{H^*(\theta)\} - H^*(\underline{\theta}))f(\theta) + \int_{\underline{\theta}}^{\overline{\theta}} \frac{dH^*(\theta)}{d\underline{\theta}} f(\theta) d\theta. \quad (3.15)$$

The first term of the right-hand side in equation (3.14) is the marginal decrease of the expected net value due to a higher $\underline{\theta}$; the integrand in (3.14) is the marginal impact of a higher $\underline{\theta}$ on the expected net value through the reduction of informational rent $\gamma(\theta)$ and the density function $f(\theta)$. Since $H^*(\underline{\theta}) > E\{H^*(\theta)\}$, it is obvious that the first term of (3.15) is negative. By the envelope theorem

$$\frac{dH^*(\theta)}{d\underline{\theta}} = \frac{\partial H^*(\theta)}{\partial \underline{\theta}} = \xi C_\theta(\mu^*(\theta), \theta) > 0.$$

Thus,

$$\xi \int_{\underline{\theta}}^{\overline{\theta}} \frac{dH^*(\theta)}{d\underline{\theta}} d\theta = \xi E\{S(\theta)\},$$

i.e., the expected informational rent that the IS manager can obtain. Since

$$H^*(\underline{\theta}) - \xi E\{S(\theta)\} > H^*(\underline{\theta}) - \xi S(\underline{\theta})$$

$$
\begin{aligned}
&> E\{NV^*(\theta)\} \\
&= E\{H^*(\theta)\},
\end{aligned}
$$

the expected organizational net value decreases as $\underline{\theta}$ increases, and the marginal effect on the expected organizational net value is the difference between the ex post organizational net value when the cost parameter is $\underline{\theta}$ and the expected organizational net value times the probability of the realization of $\underline{\theta}$.

Similarly, differentiating $E\{NV^p(\theta)\}$ with respect to $\underline{\theta}$ gives

$$
\begin{aligned}
\frac{d}{d\underline{\theta}}E\{NV^p(\theta)\} &= -NV^p(\underline{\theta})f(\theta) + \int_{\underline{\theta}}^{\overline{\theta}} NV^p(\theta)f(\theta)^2 d\theta \\
&= (E\{NV^p(\theta)\} - NV^p(\underline{\theta}))f(\theta),
\end{aligned}
$$

which is negative since $NV^p(\underline{\theta}) > E\{NV^p(\theta)\}$, and therefore $E\{NV^p(\theta)\}$ is decreasing in $\underline{\theta}$.

The Effects of Varying $\overline{\theta}$. Differentiating $E\{H^*(\theta)\}$ with respect to $\overline{\theta}$ gives

$$
\begin{aligned}
\frac{d}{d\overline{\theta}}E\{NV^*(\theta)\} &= H^*(\overline{\theta})f(\theta) + \int_{\underline{\theta}}^{\overline{\theta}} \left\{\frac{dH^*(\theta)}{d\overline{\theta}}f(\theta) - H^*(\theta)f(\theta)^2\right\} d\theta \\
&= (H^*(\overline{\theta}) - E\{H^*(\theta)\})f(\theta) + \int_{\underline{\theta}}^{\overline{\theta}} \frac{dH^*(\theta)}{d\overline{\theta}}f(\theta)d\theta
\end{aligned}
$$

Notice that when $F(\theta)$ is uniform, $H^*(\theta)$ is independent of $\overline{\theta}$, and so the last term in the above expression is zero. Since $H^*(\overline{\theta}) < E\{H^*(\theta)\}$, $E\{NV^*(\theta)\}$ is decreasing in $\overline{\theta}$.

Similarly, differentiating $E\{NV^p(\theta)\}$ with respect to $\overline{\theta}$ gives

$$
\begin{aligned}
\frac{d}{d\overline{\theta}}E\{NV^p(\theta)\} &= NV^p(\overline{\theta})f(\theta) - \int_{\underline{\theta}}^{\overline{\theta}} NV^p(\theta)f(\theta)^2 d\theta \\
&= (NV^p(\overline{\theta}) - E\{NV^p(\theta)\})f(\theta),
\end{aligned}
$$

which is clearly negative.

Thus, with a fixed $\underline{\theta}$, an increase in $\overline{\theta}$ will result in a lower expected net value for both cases. This result is obvious because both $NV^*(\theta)$ and $NV^p(\theta)$ are decreasing in θ, and a higher $\overline{\theta}$ puts positive weights on those most unfavorable states, while for all $\theta < \overline{\theta}$, $H^*(\theta)$ is unaffected by the increase in $\overline{\theta}$.

The Effects of Mean-Preserving Spreads. Since I assume that the distribution function $F(\theta)$ is uniform, the mean of the random variable is preserved if $\overline{\theta}$ is increased (decreased) and $\underline{\theta}$ is decreased (increased) by the same amount. Parameterizing the expected organizational net value by δ gives:

$$E\{H^*(\theta);\delta\} = \int_{\underline{\theta}-\delta}^{\overline{\theta}+\delta} \frac{H^*(\theta)}{\overline{\theta}-\underline{\theta}+2\delta}\, d\theta.$$

It is straightforward to show that

$$\begin{aligned}\frac{d}{d\delta}E\{H^*(\theta);\delta\}\Big|_{\delta=0} &= (H^*(\underline{\theta}) + H^*(\overline{\theta}) - \xi S(\overline{\theta}) - 2E\{H^*(\theta)\})f(\theta)\\ &= (NV^*(\underline{\theta}) + H^*(\overline{\theta}) - 2E\{NV^*(\theta)\})f(\theta).\end{aligned}$$

Notice that, since $H^*(\overline{\theta}) < NV^*(\overline{\theta})$, the effect of the distribution spread is ambiguous, and thereby the expected organizational net value may not be monotone in δ. From the previous discussion, the expected organizational net value is unambiguously decreasing in $\underline{\theta}$ and decreasing in $\overline{\theta}$. Thus, depending on which effect is stronger, the expected organizational net value can increase or decrease with respect to the spread.

Similarly,

$$\frac{d}{d\delta}E\{NV^p(\theta);\delta\}\Big|_{\delta=0} = \left(NV^p(\underline{\theta}) + NV^p(\overline{\theta}) - 2E\{NV^p(\theta)\}\right) f(\theta).$$

Since $NV^p(\theta)$ is a convex function, the above expression is positive, and thereby the expected organizational net value is monotone increasing in δ.

The Effects of Varying ξ

By the envelope theorem,

$$\begin{aligned}\frac{d}{d\xi}E\{H^*(\theta)\} &= \int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial H^*(\theta)}{\partial \xi}\, dF(\theta)\\ &= -\int_{\underline{\theta}}^{\overline{\theta}} \beta(\theta)\,\mu^*(\theta)\,dF(\theta)\\ &= -\int_{\underline{\theta}}^{\overline{\theta}} F(\theta)\,\mu^*(\theta)\,d\theta\\ &< 0,\end{aligned}$$

and the sign does not depend on any specific family of distribution functions. Since $\int_{\underline{\theta}}^{\overline{\theta}} C_\theta(\mu^*(\theta),\theta)d\theta$ is the expected marginal informational rent with respect to ξ, and

$F(\theta)$ is the probability that the realized cost parameter is less than θ, an increase in ξ decreases the expected organizational net value by an amount equal to the expected marginal informational rent.

Furthermore, for the profit center case,

$$\frac{d}{d\xi}E\{NV^p(\theta)\} = -\int_{\underline{\theta}}^{\bar{\theta}} \pi(\theta)\, dF(\theta) < 0;$$

it is just the expected profit of the profit center.

The Effects of Varying k

It is obvious that for the cost center case

$$\frac{d}{dk}E\{H^*(\theta)\} = \frac{\partial}{\partial k}E\{H^*(\theta)\} = \int_{\underline{\theta}}^{\bar{\theta}} \frac{\partial V}{\partial k}\, dF(\theta) > 0.$$

For the profit center case, let

$$NV^+ \equiv NV^p(\theta) + \xi\pi(\theta).$$

Then

$$\begin{aligned} \frac{dE\{NV^p(\theta)\}}{dk} &= \int_{\underline{\theta}}^{\bar{\theta}} \left\{ \frac{\partial NV^+}{\partial \lambda^p}\frac{d\lambda^p}{dk} + \frac{\partial NV^+}{\partial \mu^p}\frac{d\mu^p}{dk} + \frac{\partial NV^+}{\partial k} - \xi\frac{\partial \pi}{\partial k} \right\} dF(\theta) \\ &= \int_{\underline{\theta}}^{\bar{\theta}} \left\{ -\lambda^p \frac{\partial^2 V}{\partial \lambda^{p2}}\frac{d\lambda^p}{dk} + \frac{\partial V}{\partial k} - \xi\lambda^p\frac{\partial^2 V}{\partial \lambda^p \partial k} \right\} dF(\theta). \end{aligned}$$

The second equality follows from the fact that

$$\frac{\partial NV^+}{\partial \lambda^p} = -\lambda^p \frac{\partial^2 V}{\partial \lambda^{p2}},$$

that

$$\frac{\partial NV^+}{\partial \mu^p} = 0,$$

and that

$$\frac{\partial \pi}{\partial k} = \lambda^p \frac{\partial^2 V}{\partial \lambda^p \partial k}.$$

Since V is concave in λ and $V_k > 0$, the sum of the first two terms of the integrand is positive, and it reflects the increase in the organizational net value before accounting for the additional compensation to the IS department due to a higher profit, the last term

TABLE 3.1: SUMMARY OF COMPARATIVE STATICS

	k	ξ	$\underline{\theta}$	$\overline{\theta}$	δ
$E\{NV^*\}$	+	−	−	−	+/−
$E\{NV^p\}$	+	−	−	−	+

of the integrand. Thus in general the effect of an increase in k is ambiguous. If V_k is increasing, concave in λ, however, an increase in k increases the expected organizational net value, since $\xi \in [0,1]$ and

$$\frac{\partial V}{\partial k} > \lambda^p \frac{\partial^2 V}{\partial \lambda^p \partial k}.$$

The results of the comparative statics for the optimal mechanism and profit center are summarized in Table 3.1. From Table 3.1 it is clear that, except for δ, an increase in the value of parameters has an unambiguous effect on the expected organizational net value. The effect of an increase in k or ξ should be clear. A stronger demand for information processing should generate a higher organizational net value regardless of how the IS department is organized. On the other hand, the expected organizational net value should decrease if the organization's IS manager is subject to a more severe incentive problem. Since, in the presence of the IS manager's incentive problem, the case without rewards for cost savings cannot outperform the case with rewards under the optimal mechanism, organizations should get better performance from their IS departments. This hypothesis may be a topic for empirical investigation.

3.6 Benchmark: The Naive Mechanism

If the central management does not recognize the incentive problems associated with professionalism, then it may accept whatever cost standard is proposed by the IS manager and determine the effective capacity and price accordingly. Even when the central management recognizes this incentive problem and provides pecuniary rewards for cost savings, this reward must be strong enough. When the central management does not recognize the IS manager's tendency to exaggerate cost information and compensates the IS manager only with a flat salary, we call this the *naive* case. It will be instructive to compare the outcomes of the optimal mechanism I derive above with the outcomes of the naive mechanism. First I derive the effective capacity and arrival rate for the naive case.

I assume that a naive central management takes the IS manager's reported cost parameter $\hat{\theta}$ as the true cost parameter and determines the effective capacity and price that the IS department should set by solving the first-order conditions:

$$\begin{aligned} 0 &= V'(\lambda(\hat{\theta})) - W(\lambda(\hat{\theta}), \mu(\hat{\theta})) - \lambda(\hat{\theta}) W_\lambda(\lambda(\hat{\theta}), \mu(\hat{\theta})) \\ 0 &= -\lambda(\hat{\theta}) W_\mu(\lambda(\hat{\theta}), \mu(\hat{\theta})) - C_\mu(\mu(\hat{\theta}), \hat{\theta}). \end{aligned}$$

Let $\mu^n(\hat{\theta})$ and $\lambda^n(\hat{\theta})$ be the solution of the above equations. Then when $\xi > 0$, the IS manager's optimal reporting strategy can be obtained by maximizing the difference between the cost standard and the true cost, i.e., by maximizing:

$$S^n(\hat{\theta};\theta) \stackrel{\text{def}}{=} C(\mu^n(\hat{\theta}), \hat{\theta}) - C(\mu^n(\hat{\theta}), \theta)$$

with respect to $\hat{\theta}$. Differentiating $S^n(\hat{\theta};\theta)$ with respect to $\hat{\theta}$ gives

$$S^n_{\hat{\theta}}(\hat{\theta};\theta) = \left(C_\mu(\mu^n(\hat{\theta}), \hat{\theta}) - C_\mu(\mu^n(\hat{\theta}), \theta)\right) \frac{d\mu^n(\hat{\theta})}{d\hat{\theta}} + C_{\hat{\theta}}(\mu^n(\hat{\theta}), \hat{\theta}). \tag{3.16}$$

Since by assumption $C_{\hat{\theta}} > 0, C_\mu > 0$ and $C_{\mu\theta} > 0$,

$$S^n_{\hat{\theta}}(\hat{\theta};\theta)|_{\hat{\theta}=\theta} > 0,$$

the IS manager will exaggerate the cost parameter as shown in Proposition 2.2. Furthermore, it is clear that if $\mu^n(\hat{\theta})$ is non-decreasing in $\hat{\theta}$, (3.16) is always positive, and therefore the IS manager will always report the cost parameter as $\overline{\theta}$. It is thus necessary for $\mu^n(\theta)$ to be decreasing in order to have an interior solution. When this is the case, if $S^n_{\tilde{\theta}}(\tilde{\theta};\theta) > 0$ for all $\tilde{\theta} \in [\theta, \overline{\theta}]$ the IS manager still will report $\hat{\theta} = \overline{\theta}$, and if there is some $\theta^* \in [\theta, \overline{\theta}]$ such that $S^n_{\theta^*}(\theta^*;\theta) = 0$, then the IS manager will report $\hat{\theta} = \theta^*$. Let $\theta^n = \sigma^n(\theta)$ be the IS manager's optimal reporting strategy when θ is the true cost parameter of the IS department,

$$S^n(\theta^n;\theta) = C(\mu^n(\theta^n), \theta^n) - C(\mu^n(\theta^n), \theta).$$

As a result, not only does the IS manager earn an informational rent $\xi S^n(\theta)$ but both the effective capacity and arrival rate are distorted.

3.7 Imperfect Information about ξ

Given the linear structure of the IS manager's utility function, the optimal incentive scheme is linear in the IS department's cost savings: $\eta S(\theta)$, where $\eta = \xi$, when the central management has perfect information about ξ. In this section, I investigate the effect of the central management's uncertainty about ξ. When the central management is uncertain about the threshold ξ, I assume that it has some prior beliefs about the distribution of ξ. Let $Q(\xi)$ denote this probability distribution, which is assumed to be twice continuously differentiable and which has a density function $q(\xi) > 0$ if and only if $\xi \in [0, 1]$. Define

$$\begin{aligned} H^+(\lambda, \mu, \theta, \eta) &\stackrel{\text{def}}{=} V(\lambda) - \lambda W(\lambda, \mu) - C(\mu, \theta) - \eta\beta(\theta)C_\theta(\mu, \theta) \\ H^-(\lambda, \mu, \theta) &\stackrel{\text{def}}{=} V(\lambda) - \lambda W(\lambda, \mu) - C(\mu, \theta) - \beta(\theta)C_\theta(\mu, \theta). \end{aligned}$$

Then, given the linear structure of the IS manager's utility function, the IS manager will consume all the organizational slack if and only if $\eta < \xi$, so the virtual organizational net value equals $H^+(\lambda, \mu, \theta, \eta)$ if $\eta \geq \xi$, and equals $H^-(\lambda, \mu, \theta)$ otherwise. Consequently, the central management's problem becomes:

$$\max_{\lambda(\theta), \mu(\theta), \eta \in [0,1]} \int_\Theta \{Q(\eta)H^+(\lambda, \mu, \theta, \eta) + (1 - Q(\eta))\, H^-(\lambda, \mu, \theta)\}\, dF(\theta). \tag{3.17}$$

Observe that, for any μ and θ, the virtual capacity cost is:

$$C(\mu, \theta) + [1 - (1 - \eta)Q(\eta)]\beta(\theta)C_\theta(\mu, \theta).$$

Then it is clear that the optimal η can be determined solely by minimizing:

$$1 - (1 - \eta)Q(\eta). \tag{3.18}$$

Since $1 - (1 - \eta)Q(\eta) \in [0, 1]$ and reaches its maximum of 1 at the boundaries of the support of η, we must have an interior solution for η. Differentiating (3.18) and equating the expression to zero gives:

$$0 \equiv Q(\eta) - (1 - \eta)q(\eta)|_{\eta=\eta^*},$$

which implies

$$\frac{q(\eta^*)}{Q(\eta^*)} \equiv \frac{1}{1-\eta^*}. \tag{3.19}$$

To ensure that (3.18) is convex and thereby (3.19) gives a global minimum, the second-order condition is needed:

$$2q(\eta) - (1-\eta)q'(\eta) > 0, \qquad \forall \eta \in [0,1],$$

or

$$\frac{q'(\eta)}{q(\eta)} < \frac{2}{1-\eta}, \qquad \forall \eta \in [0,1]. \tag{3.20}$$

Note that $\frac{q(\eta)}{Q(\eta)}$ is the inverse hazard rate, and it is decreasing if and only if

$$\frac{q(\eta)}{Q(\eta)} > \frac{q'(\eta)}{q(\eta)}.$$

Then (3.19) yields a global minimum if $\frac{q(\eta)}{Q(\eta)}$ is monotonically decreasing, since

$$\frac{2}{1-\eta} > \frac{1}{1-\eta} = \frac{q(\eta)}{Q(\eta)} > \frac{q'(\eta)}{q(\eta)}.$$

Because $\frac{d}{d\eta}(\ln q(\eta)) = \frac{q'(\eta)}{q(\eta)}$ and $-2\frac{d}{d\eta}(\ln(1-\eta)) = \frac{2}{1-\eta}$, the second-order condition is satisfied if and only if $\ln q(\eta)$ does not increase faster than $-2\ln(1-\eta)$ does. This condition is obviously satisfied if $q(\eta)$ is decreasing or a constant, i.e., if $Q(\eta)$ is (weakly) concave. On the other hand, if the inequality in (3.20) is not satisfied globally, (3.18) can be concave over some range, and consequently the first-order condition will yield a local maximum.

Further notice that the expected η

$$\begin{aligned} \bar{\eta} &\stackrel{\text{def}}{=} \int_0^1 \eta \, dQ(\tilde{\eta}) \\ &= 1 - \int_0^1 Q(\tilde{\eta})\, d\tilde{\eta}. \end{aligned}$$

By comparison, $\eta^* \geq \bar{\eta}$ if

$$\frac{Q(\eta^*)}{q(\eta^*)} \leq \int_0^1 Q(\tilde{\eta})\, d\tilde{\eta}.$$

Assuming that the second-order condition is satisfied, substituting (3.19) into (3.18) yields

$$\alpha(\eta^*) \equiv 1 - \frac{Q(\eta^*)^2}{q(\eta^*)}. \tag{3.21}$$

For example, when $Q(\eta)$ is uniform, $\alpha(\eta^*) = 1 - \eta^{*2}$ with $\eta^* = 0.5$.

To derive the optimal mechanism, maximizing (3.17) pointwise with respect to λ and μ yields the following first-order conditions:

$$0 = V'(\lambda) - W(\lambda,\mu) - \lambda W_\lambda(\lambda,\mu) \tag{3.22}$$

$$0 = -\lambda W_\mu(\lambda,\mu) - C_\mu(\mu,\theta) - \alpha(\eta^*)\beta(\theta)C_{\theta\mu}(\mu,\theta). \tag{3.23}$$

As shown in Proposition 3.1, $\mu(\theta)$ is decreasing if

$$(1 + \alpha(\eta^*)\beta'(\theta))\, C_{\theta\mu} + \alpha(\eta^*)\beta(\theta)C_{\theta\mu\theta} > 0.$$

This condition is satisfied if both $\beta(\theta)$ and $C_{\theta\mu}(\mu,\theta)$ are non-decreasing in θ. Assume this is the case. The first-order condition (3.19) can now be given a more intuitive interpretation.

By the envelope theorem, differentiating (3.17) with respect to η totally and equating it to zero gives:

$$0 = q(\eta)\left\{\int_\Theta (H^+(\lambda,\mu,\theta,\eta) - H^-(\lambda,\mu,\theta))\, dF(\theta)\right\} - Q(\eta)\int_\Theta \frac{\partial H^+(\lambda,\mu,\theta,\eta)}{\partial\eta}\, dF(\theta). \tag{3.24}$$

From (3.24), an increase in η increases the likelihood that the IS manager will get the pecuniary reward, and thereby the expected organizational net value is increased by an amount:

$$q(\eta)\left\{\int_\Theta (H^+(\lambda,\mu,\theta,\eta) - H^-(\lambda,\mu,\theta))\, dF(\theta)\right\} = q(\eta)(1-\eta)\int_\Theta \beta(\theta)C_\theta(\mu,\theta)\, dF(\theta).$$

However, by doing so, the expected informational rent of the IS manager is also increased by an amount:

$$Q(\eta)\int_\Theta \frac{\partial H^+(\lambda,\mu,\theta,\eta)}{\partial\eta}\, dF(\theta) = Q(\eta)\int_\Theta \beta(\theta)C_\theta(\mu,\theta)\, dF(\theta).$$

Thus, at optimum, these two effects must be balanced. Furthermore, since

$$\int_\Theta \beta(\theta)C_\theta(\mu,\theta)\, dF(\theta) > 0,$$

(3.24) is equivalent to:

$$0 \equiv Q(\eta) - (1-\eta)q(\eta)|_{\eta=\eta^*},$$

as required in (3.19).

Since the central management must allocate extra budget to induce the IS manager's truthful revelation of θ under the incomplete information case, the organizational net value should be smaller than it would be if the central management had complete information. Also, the revelation principle implies that any outcomes attainable by the optimal mechanism are at least as good as the outcomes attainable by any other mechanism under the incomplete information case. Therefore the expected organizational net value attained by the optimal mechanism will be at least as large as that attained when the central management is naive. In the next section, I use a specific example to compare the results of the full-information case, the incomplete information case with the optimal mechanism, and the incomplete information case with a naive central management.

3.8 Example 3.1

The revelation principle implies that a cost center under the optimal mechanism will have a higher expected value than any other feasible IS organizational form. However, the analytical work presented thus far offers little direct insight into the magnitudes of the quantities involved. I thus present a specific example in this section to make the implications more concrete. Since this example will be used extensively in the remainder of the dissertation, this example and the related discussion are as complete as possible.

I make the following assumptions for this example:

1. The organization's information system can be characterized as an $M/M/1$ queueing system with First-Come-First-Served discipline.

2. The aggregate gross value function is $V(\lambda) = 2k\sqrt{\lambda}$, where $k > 0$, so that the gross inverse demand curve is $V'(\lambda) = \frac{k}{\sqrt{\lambda}}$. This is a member of the isoelastic family of demand curves; in this case the price elasticity of demand is $\epsilon = -2$, which is quite elastic. Thus I am modeling a user population that is fairly price sensitive, which is appropriate to situations in which the user departments have some flexibility in how they acquire computing. For example, in such a setting

a price increase on a centralized system can lead user departments to substitute their own mini- or personal-computer-based systems, while a price decrease can give rise to the implementation of new systems not previously cost effective, or substitution away from existing departmental systems.

3. The "true" capacity cost function is $C(\mu, \theta) = \theta\mu$, which implies that the marginal capacity cost is θ. Since recent empirical work (e.g., Barron [10] and Mendelson [73]) implies that constant marginal capacity cost for hardware is reasonable, a lower bound for $\underline{\theta}$ is given by the hardware component of C, which can be estimated from market data. (In Barron [10], for 1988 data, this was about $70,000 per MIPS, or about $0.002 per million instructions, assuming a 5-year system lifetime.)

4. The cost parameter $\theta \in \Theta = [\underline{\theta}, \overline{\theta}]$ is uniformly distributed. (The uniform distribution satisfies the monotonicity constraint.)

5. As has been assumed throughout, $\xi \in [0, 1]$.

6. The users are assumed to have a linear delay cost function of v dollars per unit of time, i.e., $W(\lambda, \mu) = vE\{\tilde{W}(\lambda, \mu)\} = \frac{v}{\mu - \lambda}$. There is no further loss of generality in doing so, but this implies that k and θ are correspondingly scaled by v.

I consider several cases in what follows. Naturally, I derive and discuss the optimal mechanism. In addition, in order to have some benchmarks as to whether or not the mechanism is worth the effort, I also provide results for the perfect information case, which of course is not in general attainable (however, see the "Discussion" below), and two cases which are attainable, the profit center and the naive mechanism. Under the naive mechanism the central management accepts whatever report the IS manager gives and sets capacity and arrival rate accordingly. This corresponds to an extreme case of bounded rationality (Simon [100]; Williamson [116]). Of course the revelation principle guarantees that the expected net value of the optimal mechanism will exceed that of either the profit center or the naive mechanism, but this result is gross of any costs of computation and communication. Thus if either of these suboptimal alternatives is not too inferior to the optimal mechanism, they could in fact be superior once these additional costs are included.

The results for this example are summarized in Table 3.2, and numerical results for a particular case are shown in Table 3.3. More detailed discussion follows.

Incomplete Information with Optimal Mechanism

It is easy to see that the optimal mechanism given in Proposition 3.1 is feasible for this example because $C_{\theta\mu\theta} = 0$ and $\beta'(\theta) > 0$. It is straightforward to show that the Hessian matrix of the pointwise virtual organizational net value H is negative definite if, for all θ in Θ, $k > \sqrt{\gamma(\theta)}$, where $\gamma(\theta) = \theta + \xi\beta(\theta) = \theta + \xi(\theta - \underline{\theta})$ is the virtual marginal capacity cost. Assume this is the case, then from (3.11) and (3.12),

$$\begin{aligned} p^*(\theta) &= \gamma(\theta) \\ \lambda^*(\theta) &= \left[\frac{k - \sqrt{\gamma(\theta)}}{\gamma(\theta)}\right]^2 \\ \mu^*(\theta) &= \frac{k(k - \sqrt{\gamma(\theta)})}{\gamma(\theta)^2}. \end{aligned}$$

It is clear that $\lambda^*(\theta)$ and $\mu^*(\theta)$ are positive if $k > \sqrt{\gamma(\theta)}$, which just is the sufficient condition for the central management's problem to yield a unique maximum for every θ in Θ. Since $\gamma(\theta)$ is increasing in θ and $\mu^*(\theta)$ is decreasing in $\gamma(\theta)$ whenever $k > \sqrt{\gamma(\theta)}$, $\mu^*(\theta)$ is decreasing in θ, and thereby the monotonicity constraint is satisfied.

Let $T^*(\theta)$ denote the optimal lump-sum budget allocation to the IS department; then

$$T^*(\theta) = \theta\mu^*(\theta) + \int_\theta^{\overline{\theta}} \mu^*(\tilde{\theta})d\tilde{\theta}.$$

The informational rent of the IS manager therefore equals $\xi \int_\theta^{\overline{\theta}} \mu^*(\tilde{\theta})d\tilde{\theta}$. Obviously, as asserted by Lemma 3.1, this informational rent is strictly decreasing in θ and equals zero when $\theta = \overline{\theta}$. Direct calculation shows that the expected organizational net value:

$$\begin{aligned} E\{NV^*(\theta)\} &= E\{V(\lambda^*(\theta)) - \lambda^*(\theta)W(\lambda^*(\theta),\mu^*(\theta)) - C(\mu^*(\theta),\theta) - \xi S^*(\theta)\} \\ &= E\{V(\lambda^*(\theta)) - \lambda^*(\theta)W(\lambda^*(\theta),\mu^*(\theta)) - C(\mu^*(\theta),\theta) - \xi\beta(\theta)C_\theta(\mu^*(\theta),\theta)\} \\ &= \int_\Theta \gamma(\theta)\lambda^*(\theta)dF(\theta). \end{aligned}$$

Note that the optimal utilization rate $\rho^*(\theta) \equiv \frac{\lambda^*(\theta)}{\mu^*(\theta)} = \frac{k-\sqrt{\gamma(\theta)}}{k}$ is decreasing in θ, while the optimal waiting time $W^*(\theta) \equiv \frac{1}{\mu^*(\theta)-\lambda^*(\theta)} = \frac{\gamma(\theta)^{3/2}}{k-\sqrt{\gamma(\theta)}}$ is increasing in θ. So when the IS department is more efficient, the system response time should be shorter, even though the system is running with a heavier load. Nevertheless the aggregate users' delay $\lambda^*(\theta)W^*(\theta) = \frac{k-\sqrt{\gamma(\theta)}}{\sqrt{\gamma(\theta)}}$ is decreasing in θ. In fact, under a fairly general condition, the optimal utilization rate is increasing in λ, while the optimal waiting time is decreasing in λ. Consequently, a large organization should have a higher utilization rate while having a faster response time in general. These hypotheses should be worth some further empirical investigation.

Upper Benchmark: A Fully Informed Central Management

It is clear that the optimal solution of the full-information case can be obtained by replacing $\gamma(\theta)$ by θ. Then

$$\begin{aligned} \lambda^f(\theta) &= \left[\frac{k-\sqrt{\theta}}{\theta}\right]^2 \\ \mu^f(\theta) &= \frac{k(k-\sqrt{\theta})}{\theta^2} \\ p^f(\theta) &= \theta \end{aligned}$$

and the expected organizational net value:

$$E\{NV^f(\theta)\} = \int_\Theta \theta\lambda^f(\theta)dF(\theta).$$

When the central management is fully informed, the optimal subsidy to the IS department $\theta(\mu^f(\theta) - \lambda^f(\theta)) = \frac{k-\sqrt{\theta}}{\sqrt{\theta}}$, which is equal to the aggregate users' delay cost, $\frac{\lambda^f(\theta)}{\mu^f(\theta)-\lambda^f(\theta)}$. This conforms to the usual optimality condition for an $M/M/1$ queueing system with linear user delay cost and capacity cost (Dewan and Mendelson [25]).

Discussion. The effects of the information asymmetry can now be clearly seen. For every θ, $NV^f(\theta) > NV^*(\theta)$ and

$$\lambda^*(\theta) \le \lambda^f(\theta);\ \mu^*(\theta) \le \mu^f(\theta);\ p^*(\theta) \ge p^f(\theta);\ \rho^*(\theta) \le \rho^f(\theta);\ W^*(\theta) \ge W^f(\theta).$$

Since $\gamma(\theta) \to \theta$ as $\theta \to \underline{\theta}$,

$$\lambda^*(\theta) \to \lambda^f(\theta);\ \mu^*(\theta) \to \mu^f(\theta);\ p^*(\theta) \to p^f(\theta);\ \rho^*(\theta) \to \rho^f(\theta);\ W^*(\theta) \to W^f(\theta)$$

as $\theta \to \underline{\theta}$. As mentioned before, this is the "no distortion at the bottom" property.

Thus, when $\theta = \underline{\theta}$, the system is operating at its full efficiency; however, the ex post informational rent of the IS department is also the largest. On the other hand, when $\theta = \overline{\theta}$, the state of the system is set by taking full account of the virtual informational rent, $\xi(\overline{\theta} - \underline{\theta})\mu^*(\overline{\theta})$, but the IS manager's ex post informational rent equals zero. The intuition here is that the central management attempts to distort the capacity decision so that the informational rent can be reduced when θ is low. Hence, even when the IS department has no way to overstate its cost (i.e., $\theta = \overline{\theta}$), the capacity is set as if the marginal cost $\gamma(\overline{\theta}) = \overline{\theta} + \xi(\overline{\theta} - \underline{\theta})$. By distorting the capacity decision, the central management is able to make overstating the cost parameter relatively unattractive, and thereby reduce the IS department's incentive for misrepresentation.

Note that $(\lambda^f(\theta), \mu^f(\theta))$ is feasible even in the presence of asymmetric information because $\mu^f(\theta)$ is decreasing in θ. However, the mechanism implementing the full information outcomes is not optimal because an excessive budget is required to induce truth-revelation. In order to induce the IS manager's truth-revelation while implementing $(\lambda^f(\theta), \mu^f(\theta))$, the central management must set:

$$S(\theta) = \int_{\theta}^{\overline{\theta}} C_\theta(\mu^f(\tilde{\theta}), \tilde{\theta})d\tilde{\theta},$$

which will give the IS manager too large an informational rent and leave the organization worse off.

The Profit Center

As for the cost center case, the central management must provide some incentive in order to induce the IS manager not to consume her department's profit. From Section 3.2 the optimal incentive scheme is $\xi\pi$, so that $U(\theta) = \xi\pi(\theta)$, where

$$\pi(\theta) = \max_{\lambda,\mu} \lambda(V'(\lambda) - W(\lambda, \mu)) - C(\mu, \theta).$$

It can be easily verified that for the current example the above program has a unique maximum if $k > 2\sqrt{\theta}$. Assuming this is the case gives

$$\lambda^p(\theta) = \left[\frac{k - 2\sqrt{\theta}}{2\theta}\right]^2$$

$$\mu^p(\theta) = \frac{k\,(k-2\sqrt{\theta}\,)}{4\theta^2}$$
$$p^p(\theta) = \frac{2\,\theta\,(k-\sqrt{\theta}\,)}{k-2\sqrt{\theta}}.$$

The monopolistic price $p^p(\theta)$ is thus more than double the net value maximizing price θ even though $p^p(\theta)$ is determined based on the true cost. Also note that the IS department's revenue

$$\begin{aligned} p^p(\theta)\lambda^p(\theta) &= \theta\left[1+\frac{k}{k-2\sqrt{\theta}}\right]\lambda^p(\theta) \\ &= \theta\left[1+\frac{\mu^p(\theta)}{\lambda^p(\theta)}\right]\lambda^p(\theta) \\ &= \theta\,(\lambda^p(\theta)+\mu^p(\theta)). \end{aligned}$$

So the IS department will make a profit equal to $\theta\lambda^p(\theta)$, since its capacity cost is $\theta\mu^p(\theta)$. Consequently, the expected organizational net value equals:

$$\begin{aligned} E\{NV^p(\theta)\} &= E\{V(\lambda^p(\theta)) - \lambda^p(\theta)\,V'(\lambda^p) + (1-\xi)\,\pi(\theta)\} \\ &= \int_\Theta \left\{k\sqrt{\lambda^p(\theta)} + (1-\xi)\,\theta\,\lambda^p(\theta)\right\} dF(\theta). \end{aligned}$$

Furthermore, the utilization rate of the system under the profit center $\rho^p(\theta) = \frac{k-2\sqrt{\theta}}{k}$. Since $\xi \in [0,1]$, and thereby $\sqrt{\gamma(\theta)} = \sqrt{(1+\xi)\theta - \xi\underline{\theta}} < 2\sqrt{\theta}$, the profit center will have a lower utilization rate than the cost center for all θ.

Whether the mean waiting time is larger or smaller under the profit center is less obvious, however. Since $W^p(\theta) = \frac{2\theta\sqrt{\theta}}{k-2\sqrt{\theta}}$ and $2\sqrt{\theta} > \sqrt{\gamma(\theta)}$, a sufficient condition for $W^p(\theta) > W^*(\theta)$ is:

$$\underline{\theta} > \theta\left[1+\frac{1-4^{1/3}}{\xi}\right]. \tag{3.25}$$

Thus, as ξ approaches zero, the mean waiting time under the profit center will be larger. Clearly, if $\underline{\theta} > \overline{\theta}\left[2-4^{1/3}\right]$, $W^p(\theta) > W^*(\theta)$ for all $\theta \in \Theta$ and $\xi \in [0,1]$. In other words, as long as the support of θ is not too wide (i.e., the central management is not too uncertain about the IS department's cost), the users will enjoy a shorter delay on average when the IS department is organized as a cost center. (Thus $W^f(\theta) < W^p(\theta)$ for all θ.) The intuition is that if Θ is very wide and ξ is close to one, the distortion will be significant when θ is large. Consequently, the mean waiting time under the cost center may become larger than that under the profit center. In general the users will be charged a higher price as well as suffer a longer delay when the IS department is organized as a profit center. In any case, because $V'(\lambda^p(\theta)) > V'(\lambda^*(\theta))$, the *total cost per job* incurred by the users under a profit center will be larger than under a cost center. Consequently, the users will press the central management to reorganize the IS department as a cost center. It is then not surprising to see that the majority of firms organize their IS departments as cost centers (McGee [69]).

Lower Benchmark: A Naive Central Management

First the IS manager's optimal reporting strategy must be determined. Let $\hat{\theta}$ be the cost parameter reported by the IS manager. As in the previous case, $\hat{\theta}$ can replace θ to give:

$$\begin{aligned} \mu^n(\hat{\theta}) &= \frac{k(k-\sqrt{\hat{\theta}})}{\hat{\theta}^2} \\ S^n(\hat{\theta};\theta) &= (\hat{\theta}-\theta)\mu^n(\hat{\theta}). \end{aligned}$$

The effects of failing to induce the IS manager to report truthfully can now be seen clearly. Since

$$S^n_{\hat{\theta}}(\hat{\theta};\theta) = (\hat{\theta}-\theta)\frac{d\mu^n(\hat{\theta})}{d\hat{\theta}} + \mu^n(\hat{\theta}),$$

and $S^n_{\hat{\theta}}(\hat{\theta};\theta)|_{\hat{\theta}=\theta} = \mu^n(\theta) > 0$, the optimality of the manager's reporting strategy requires

$$S^n_{\hat{\theta}}(\hat{\theta};\theta) \geq 0.$$

Thus, if $S^n_{\hat{\theta}}(\hat{\theta};\theta) > 0$ for all $\hat{\theta} \in [\underline{\theta}, \overline{\theta}]$, the IS manager will always report $\hat{\theta} = \overline{\theta}$. In order for $S^n(\hat{\theta},\theta)$ to reach a maximum within Θ, the second-order necessary condition requires $S^n_{\hat{\theta}}(\hat{\theta};\theta)$ to be concave in $\hat{\theta}$. In other words,

$$S^n_{\hat{\theta}\hat{\theta}}(\hat{\theta};\theta) = (\hat{\theta}-\theta)\frac{d^2\mu^n(\hat{\theta})}{d\hat{\theta}^2} + 2\frac{d\mu^n(\hat{\theta})}{d\hat{\theta}} \leq 0.$$

A sufficient condition for $S^n(\hat{\theta},\theta)$ to be strictly concave in $\hat{\theta}$ is $3\underline{\theta} - \overline{\theta} > 0$ and

$$\frac{3}{8k} < \frac{1}{\sqrt{\theta}}\left[\frac{3\theta-\overline{\theta}}{5\theta-\overline{\theta}}\right].$$

So when the central management's uncertainty about the cost parameter is not too severe (i.e., the support of θ is not too wide) and/or the demand parameter k is sufficiently large, $S^n(\hat\theta;\theta)$ is globally concave in $\hat\theta$. Thus, if $S^n_{\hat\theta}(\hat\theta;\theta)|_{\hat\theta=\overline{\theta}} < 0$, there is an interior solution, and the IS manager's optimal reporting strategy is the solution to the first-order condition:

$$S^n_{\hat\theta}(\hat\theta;\theta) = (\hat\theta - \theta)\frac{d\mu^n(\hat\theta)}{d\hat\theta} + \mu^n(\hat\theta) = 0.$$

Writing out the second equality and rearranging the terms yields:

$$\begin{aligned}\theta &= \hat\theta\left[1 - \frac{2k - 2\sqrt{\hat\theta}}{4k - 3\sqrt{\hat\theta}}\right] \\ &\equiv g(\hat\theta). \end{aligned} \tag{3.26}$$

It is easy to see that, for $k > \sqrt{\hat\theta}$, $g(\hat\theta)$ is increasing and $g(\theta) < \theta$. Thus $q(\hat\theta)$ is invertible, and, for all $\theta > 0$, the first-order condition (3.26) has a unique solution given by the inverse function of $g(\hat\theta)$: $\hat\theta(\theta) \stackrel{\text{def}}{=} g^{-1}(\theta)$. Thus, if $\hat\theta(\theta) \geq \overline{\theta}$ for all $\theta \in \Theta$, the IS manager will always report $\hat\theta(\theta) = \overline{\theta}$; if there is a $\tilde\theta \in (\underline{\theta}, \overline{\theta})$ such that $\hat\theta(\tilde\theta) = \overline{\theta}$, the IS manager's optimal reporting strategy is

$$\theta^n(\theta) = \begin{cases} \overline{\theta} & \text{if } \theta \geq \tilde\theta \\ \hat\theta & \text{where } g(\hat\theta) = \theta \text{ for } \theta \in [\underline{\theta}, \tilde\theta). \end{cases}$$

The following lemma shows that the IS manager tends to exaggerate the cost more when her department's cost is high.

LEMMA 3.2 *For Example 3.1, when $\mu^n(\hat\theta) > 0$, i.e., $k > \sqrt{\hat\theta}$, $g(\hat\theta)$ is increasing and the mapping of the IS manager's report $g^{-1}(\theta) \equiv \hat\theta(\theta) > \theta$. Furthermore, the IS manager tends to exaggerate more when θ is high.*

PROOF. Given $k > \sqrt{\hat\theta}$, $g(\hat\theta) < \hat\theta$. It is straightforward to show that

$$\frac{dg(\hat\theta)}{d\hat\theta} = 1 - \frac{6\,(2k - \sqrt{\hat\theta})(k - \sqrt{\hat\theta})}{(4k - 3\sqrt{\hat\theta})} \in (0,1).$$

Since $g(\hat\theta)$ is monotone, the inverse mapping $g^{-1}(\theta)$ is well-defined and $g^{-1}(\theta) = \hat\theta(\theta) > \theta$. To prove the last part of the lemma, it suffices to show that

$$\frac{d}{d\theta}\left(\hat\theta(\theta) - \theta\right) > 0,$$

or simply $\frac{d\hat{\theta}(\theta)}{d\theta} > 1$. By the inverse function theorem,

$$\left.\frac{dg^{-1}(\theta)}{d\theta}\right|_{\theta^o} = \left.\left[\frac{dg(\hat{\theta})}{d\hat{\theta}}\right]^{-1}\right|_{\hat{\theta}=g^{-1}(\theta^o)}.$$

But $\frac{dg(\hat{\theta})}{d\hat{\theta}} \in (0,1)$, proving the lemma. ||

As for the optimal mechanism:

$$\begin{aligned} \lambda^n(\theta) &= \left[\frac{k-\sqrt{\theta^n}}{\theta^n}\right]^2 \\ \mu^n(\theta) &= \frac{k\,(k-\sqrt{\theta^n})}{\theta^{n\,2}} \\ p^n(\theta) &= \theta^n, \end{aligned}$$

and the organization's expected net value equals:

$$E\{NV^n(\theta)\} = \int_{\Theta} \lambda^n(\theta^n)\theta^n dF(\theta).$$

So, for each θ, $\lambda^n(\theta)$, and thereby $NV^n(\theta)$, is positive if and only if $k - \sqrt{\theta^n} > 0$. The extra allocation consumed by the IS manager equals $S^n(\theta) = C(\mu^n(\theta^n), \theta^n) - C(\mu^n(\theta^n), \theta)$, and, since the IS manager will consume all the excess budget allocation, $S^n(\theta)$, this is also the informational cost to the central management.

Numerical Results

To complete this example, I provide numerical results for the case where $k = 5$, $\xi = 0.5$ and $\Theta = [1,2]$ in Table 3.3 and Figures 3.1 to 3.8. Since $\xi \in [0,1]$, this value for ξ corresponds to a moderate level of incentive conflict. Since the functional forms used here imply that θ is the marginal capacity cost (MCC), and θ is measured in units of the users' delay cost parameter, v, $\theta = 1$ implies that the central management's lower bound on the MCC equals v. For the sake of reference, this lower bound must be at least as large as the hardware component of MCC, so the value mentioned in Assumption 3 at the start of the example corresponds roughly to $v = \$7.50$ per hour, which we can take to be approximately equal to the average user's wage rate. Thus $\theta = 1$ is consistent with any wage rate of at least \$7.50 per hour, which would generally be the case. The value of k is also measured in units of v, but it is not quite as easy to interpret as θ since the marginal value of jobs depends on λ. The maximum λ in Table 3.3 implies that the gross value of computing is \$40v per period, so if v = \$15 per hour, this is about \$1.25 million per year, a fairly modest amount. As a result, the numerical values used model a modest computing facility.

Since $\Theta = [1,2]$ satisfies (3.25) for all θ:

$$\begin{aligned} \lambda^f(\theta) &\geq \lambda^*(\theta) > \lambda^p(\theta) \\ \mu^f(\theta) &\geq \mu^*(\theta) > \mu^p(\theta) \\ p^f(\theta) &\leq p^*(\theta) < p^p(\theta) \\ \rho^f(\theta) &\geq \rho^*(\theta) > \rho^p(\theta) \\ W^f(\theta) &\leq W^*(\theta) < W^p(\theta). \end{aligned}$$

From Figure 3.5, utilization rates are decreasing in θ, but Figure 3.4 nevertheless shows that waiting times are increasing. As can be seen in Figure 3.2, this stems from the rapid decreases in optimal capacity as θ increases. Similarly, prices are increasing in θ as shown in Figure 3.3, and therefore users suffer both higher prices and longer delays as costs of low IS efficiency. Furthermore, the decreases in capacity and utilization rate taken together imply the falling arrival rate in Figure 3.1, increasing the opportunity costs of jobs not served as θ rises, thereby adding further to the costs of low efficiency. Note that in all of the figures the profit center almost always performs worse than the other alternatives.

Figure 3.8 shows that in this example the ex post organizational net value is higher under the optimal mechanism than under the profit center for all values of θ. (Note that the revelation principle implies that the expected net value under the optimal mechanism must be higher than for the profit center, but not necessarily that the ex post net value be higher for all values of θ.) Interestingly, and perhaps surprisingly, Figure 3.7 shows that over most of θ's range the IS manager's utility is also significantly higher under the optimal mechanism than it is under the profit center; the IS manager will prefer a profit center only when she runs a high-cost operation. Clearly the restriction in capacity (and output) under the profit center is severe enough to make her informational rent very low, and this restriction has a similarly bad effect on the organization as a whole. This is strikingly apparent in the *TNV* panel of Table 3.3, which shows that the profit center yields a much smaller total "pie," which adversely affects all participants. Conversely, the optimal mechanism yields *TNV*s that are a very large fraction of the full-information case. The vastly inferior performance of the profit center indicates that a smaller informational rent for the IS manager is not necessarily organizationally beneficial. Also note from Figure 3.7 that the amounts paid to the IS manager under the optimal mechanism are not extreme, lying between zero and about $117,000 per year if v = $15 per hour, and amounting to about $34,000 per year at the expected θ of 1.5.

The preceding discussion implies that a profit center will generally be disliked by all three major participants. It must be remembered, however, that there are two important assumptions behind this conclusion: (1) no external market access, and (2) unlimited communication between the IS manager and the central management.

When the central management is uncertain about ξ, given the assumptions, it is easy to verify that (3.22)–(3.24) yield a global optimal solution. Thus, by setting $\gamma(\theta) = \theta + \alpha(\eta^*)(\theta - \underline{\theta})$, the expected organizational net value equals:

$$\begin{aligned} E\{NV(\theta)\} &= \int_{\Theta} \gamma(\theta)\lambda^*(\theta)dF(\theta) \\ &= \left[\frac{\theta}{\overline{\theta}-\underline{\theta}} + \frac{k^2 \ln \gamma(\theta)}{(\overline{\theta}-\underline{\theta})(1+\alpha(\eta^*))} - \frac{4k\sqrt{\gamma(\theta)}}{(\overline{\theta}-\underline{\theta})(1+\alpha(\eta^*))}\right]_{\underline{\theta}}^{\overline{\theta}}. \end{aligned}$$

If $Q(\eta)$ is uniform and $F(\theta)$ is uniform over [1, 2], then $E\{NV(\theta)\}$ is approximately equal to 7.93. Compared with knowing ξ exactly, the expected organizational net value decreases about 0.6 when the central management is uncertain about ξ and the realization of ξ equals 0.5.

For the profit center case, to determine the optimal incentive scheme, the central management solves

$$\max_{\eta \in [0,1]} \int_{\Theta} \{Q(\eta)NV^+(\lambda^p(\theta), \mu^p(\theta), \theta, \eta) + (1 - Q(\eta))\, NV^-(\lambda^p(\theta), \mu^p(\theta), \theta)\}\, dF(\theta), \tag{3.27}$$

where $\lambda^p(\theta)$ and $\mu^p(\theta)$ are the IS department's profit-maximizing solution and

$$\begin{aligned} NV^+(\lambda^p(\theta),\mu^p(\theta),\theta,\eta) &= V(\lambda^p(\theta)) - \lambda^p(\theta)W(\lambda^p(\theta),\mu^p(\theta)) - C(\mu^p(\theta),\theta) - \eta\pi(\theta) \\ NV^-(\lambda^p(\theta),\mu^p(\theta),\theta) &= V(\lambda^p(\theta)) - \lambda^p(\theta)W(\lambda^p(\theta),\mu^p(\theta)) - C(\mu^p(\theta),\theta) - \pi(\theta). \end{aligned}$$

It is easy to see that (3.27) yields the same first-order condition with respect to η as in the cost center case:

$$0 \equiv q(\eta)(1-\eta) - Q(\eta)|_{\eta=\eta^*}.$$

For the profit center, therefore,

$$E\{NV^p\} = \int_\Theta \left\{ k\sqrt{\lambda^p(\theta)} + (1-\alpha(\eta^*))\theta\lambda^p(\theta) \right\} dF(\theta),$$

and when $Q(\xi)$ is uniform, $E\{NV^p\}$ is approximately equal to 4.82, versus 5.12 when ξ is known exactly.

As for the comparative statics, I show numerically in Table 3.4 and Figure 3.9 that $E\{H^*(\theta)\}$ is decreasing in $\underline{\theta}$ over [0.1, 2.0], and thereby $\frac{d}{d\underline{\theta}}E\{H^*(\theta)\}$ is negative over the same range of $\underline{\theta}$. Intuitively, even though a higher $\underline{\theta}$ results in a lower virtual cost $\gamma(\theta)$ for every $\theta > \underline{\theta}$, the reduction in the net value, $-NV^*(\underline{\theta})$, more than offsets the effects of the reduction in the virtual costs. The numerical results for $E\{NV^p(\theta)\}$ with $\underline{\theta}$ varying from 0.1 to 2.0 also can be found in Table 3.4 and in Figure 3.9. Regardless of whether the IS department is organized as a cost center (with the optimal mechanism) or as a profit center, it is obvious that the expected net value is a decreasing function of $\underline{\theta}$ in both cases and that the cost center outperforms the profit center over the entire range.

Table 3.5 and Figure 3.10 contain the results for both $E\{NV^*(\theta)\}$ and $E\{NV^p(\theta)\}$ with $\overline{\theta}$ ranging from 1.1 to 3.0 and $\underline{\theta}$ fixed at 1. From Figure 3.10, $E\{NV^*(\theta)\}$ is clearly higher than $E\{NV^p(\theta)\}$ for all $\overline{\theta} \in [1.1, 3.0]$, although the difference decreases as $\overline{\theta}$ increases. Table 3.6 and Figure 3.11 contain the numerical results for $E\{NV^*(\theta)\}$ and $E\{NV^p(\theta)\}$ with δ varying over $[-0.4, 0.9]$. Figure 3.11 shows that when the IS department is organized as a profit center, the expected net value is increasing, convex in δ, and when δ ranges from -0.4 to 0.4, the expected net value is close to a flat line. That is, the expected net value is not very sensitive to the spread or shrinkage of the support of the distribution over the range [0.6, 2.4] with the mean fixed at 1.5. This insensitivity to the distribution of the cost parameter also holds true when the IS department is organized as a cost center under the optimal mechanism.

Table 3.7 and Figure 3.12 give the results for $E\{NV^*(\theta)\}$ and $E\{NV^p(\theta)\}$ with ξ varying from 0 to 1. In both cases, the effects of varying ξ on the expected net values are moderate and the expected net values are approximately linear in ξ. When ξ increases from 0 to 1, the expected net value decreases by 2.63, approximately a 26 percent decrease, for the cost center case, and by 1.19, approximately a 21 percent decrease, for the profit center case. Again the cost center outperforms the profit center for all $\xi \in [0, 1]$, but the difference decreases from 4.33 to 2.89 as ξ increases from 0 to 1. The positive effect of an increase in k on both cases is clear, as shown in Table 3.8 and Figure 3.13.

TABLE 3.2: EXAMPLE 3.1—SUMMARY OF THE FOUR CASES CONSIDERED.

$p(\theta)$	$p^*(\theta) = \gamma(\theta) = \theta + \xi(\theta - \underline{\theta})$ $p^f(\theta) = \theta$ $p^p(\theta) = \frac{2\theta(k-\sqrt{\theta})}{k-2\sqrt{\theta}}$ $p^n(\theta) = \theta^n$
$\lambda(\theta)$	$\lambda^*(\theta) = \left[\frac{k-\sqrt{\gamma(\theta)}}{\gamma(\theta)}\right]^2$ $\lambda^f(\theta) = \left[\frac{k-\sqrt{\theta}}{\theta}\right]^2$ $\lambda^p(\theta) = \left[\frac{k-2\sqrt{\theta}}{2\theta}\right]^2$ $\lambda^n(\theta) = \left[\frac{k-\sqrt{\theta^n}}{\theta^n}\right]^2$
$\mu(\theta)$	$\mu^*(\theta) = \frac{k\,(k-\sqrt{\gamma(\theta)})}{\gamma(\theta)^2}$ $\mu^f(\theta) = \frac{k\,(k-\sqrt{\theta})}{\theta^2}$ $\mu^p(\theta) = \frac{k\,(k-\sqrt{2\theta})}{4\theta^2}$ $\mu^n(\theta) = \frac{k\,(k-\sqrt{\theta^n})}{\theta^{n\,2}}$
$E\{NV(\theta)\}$	$E\{NV^*\} = \int_\Theta \gamma(\theta)\lambda^*(\theta)dF(\theta)$ $E\{NV^f\} = \int_\Theta \theta\lambda^f(\theta)dF(\theta)$ $E\{NV^p\} = \int_\Theta \left\{k\sqrt{\lambda^p(\theta)} + (1-\xi)\theta\lambda^p(\theta)\right\} dF(\theta)$ $E\{NV^n\} = \int_\Theta \theta^n\lambda^n(\theta)dF(\theta)$

TABLE 3.3: EXAMPLE 3.1—NUMERICAL RESULTS.

θ	λ^f	λ^*	λ^n	λ^p	μ^f	μ^*	μ^n	μ^p	p^f	p^*	p^n	p^p
1.0	16.00	16.00	3.91	2.25	20.00	20.00	5.36	3.75	1.00	1.00	1.84	2.67
1.2	10.59	8.82	3.21	1.37	13.56	11.41	4.48	2.44	1.20	1.30	2.00	3.25
1.4	7.43	5.45	3.21	0.88	9.74	7.30	4.48	1.68	1.40	1.60	2.00	3.83
1.6	5.45	3.63	3.21	0.60	7.30	5.02	4.48	1.21	1.60	1.90	2.00	4.40
1.8	4.13	2.56	3.21	0.41	5.65	3.63	4.48	0.89	1.80	2.00	2.00	4.97
2.0	3.21	1.87	3.21	0.29	4.48	2.74	4.48	0.68	2.00	2.50	2.00	5.53
θ	W^f	W^*	W^n	W^p	ρ^f	ρ^*	ρ^n	ρ^p	S^f	S^*	S^n	π^p
1.0	0.25	0.25	0.69	0.67	0.80	0.80	0.73	0.60	0.00	7.55	4.52	2.25
1.2	0.33	0.38	0.79	0.94	0.78	0.77	0.72	0.56	0.00	4.52	3.59	1.64
1.4	0.43	0.54	0.79	1.26	0.76	0.75	0.72	0.53	0.00	2.70	2.69	1.24
1.6	0.54	0.72	0.79	1.64	0.75	0.72	0.72	0.49	0.00	1.49	1.79	0.95
1.8	0.66	0.93	0.79	2.08	0.73	0.70	0.72	0.46	0.00	0.63	0.90	0.75
2.0	0.79	1.16	0.79	2.60	0.72	0.68	0.72	0.43	0.00	0.00	0.00	0.59
θ	U^f	U^*	U^n	U^p	NV^f	NV^*	NV^n	NV^p	TNV^f	TNV^*	TNV^n	TNV^p
1.0	0.00	3.78	2.26	1.13	16.00	12.22	7.20	8.63	16.00	16.00	11.72	9.76
1.2	0.00	2.26	1.79	0.82	12.70	10.34	6.43	6.67	12.70	12.60	10.01	7.49
1.4	0.00	1.35	1.34	0.62	10.41	8.83	6.43	5.32	10.41	10.18	9.12	5.94
1.6	0.00	0.74	0.90	0.48	8.72	7.66	6.43	4.34	8.72	8.40	8.22	4.82
1.8	0.00	0.31	0.45	0.37	7.44	6.76	6.43	3.59	7.44	7.07	7.33	3.96
2.0	0.00	0.00	0.00	0.29	6.43	6.04	6.43	3.01	6.43	6.04	6.43	3.30

$E\{NV^f\} = 10.01$ (Full Information)

$E\{NV^*\} = 8.52$ (Optimal Mechanism)

$E\{NV^n\} = 6.45$ (Naive)

$E\{NV^p\} = 5.12$ (Profit Center)

Here ρ's are the utilization rates; TNV's are the maximum total net values that can be generated by the system, where $TNV^f = U^f + NV^f, TNV^* = U^* + NV^*, TNV^n = S^n + NV^n, and TNV^p = U^p + NV^p$. Note that in the naive case the central management loses a full amount of S^n as the informational rent, so TNV^n equals $S^n + NV^n$ instead of $U^n + NV^n$. Also notice that, for the naive case, $\tilde{\theta} \approx 1.09$, and so $\theta^n = \overline{\theta}$ for all θ larger than 1.09. The quantities p, S, U, NV, and TNV are all measured in terms of v, the delay cost parameter, so they can be translated to dollars per period for any appropriate v. The quantities λ and μ most typically would be measured in MIPS or transactions per second.

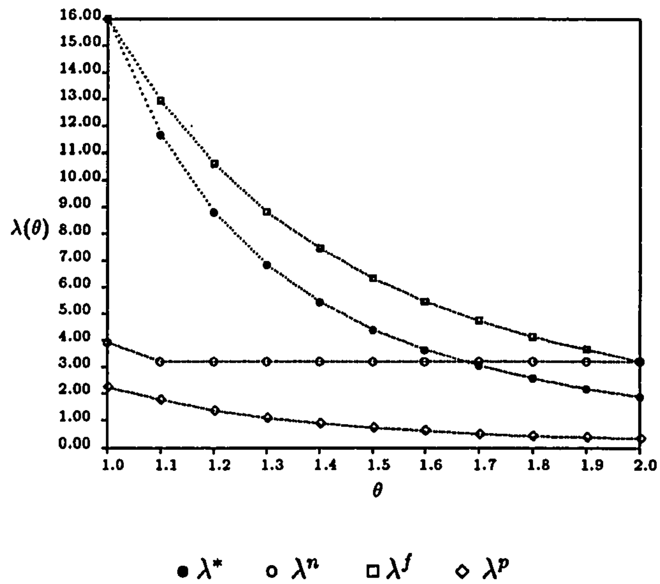


FIGURE 3.1: EXAMPLE 3.1—$\lambda(\theta)$.

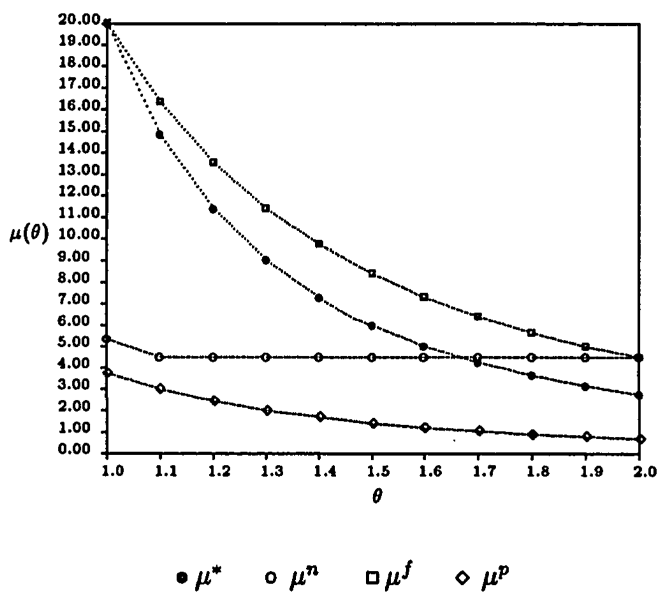


FIGURE 3.2: EXAMPLE 3.1—$\mu(\theta)$.

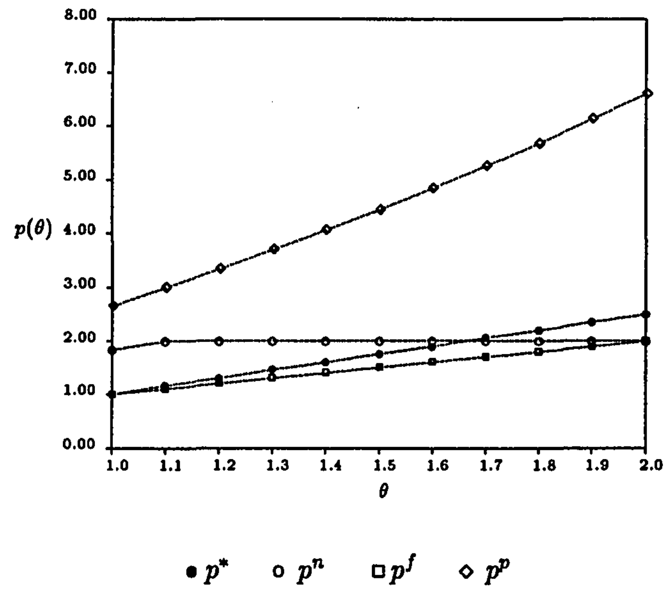


FIGURE 3.3: EXAMPLE 3.1—$p(\theta)$.

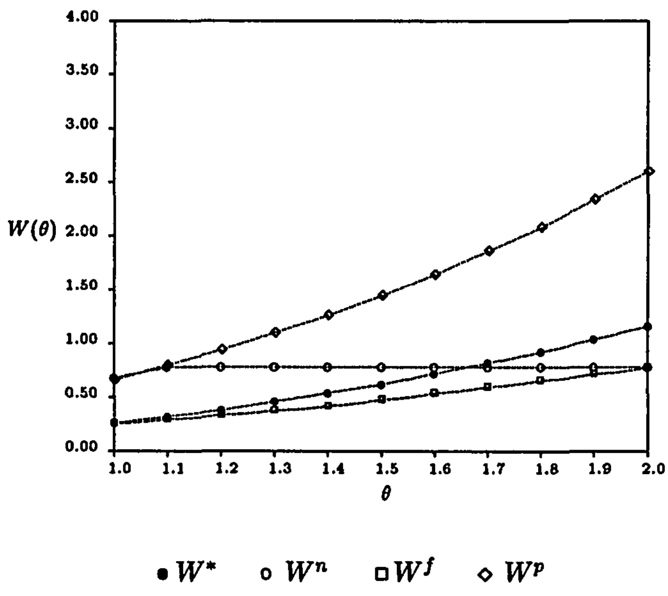


FIGURE 3.4: EXAMPLE 3.1—$W(\theta)$.

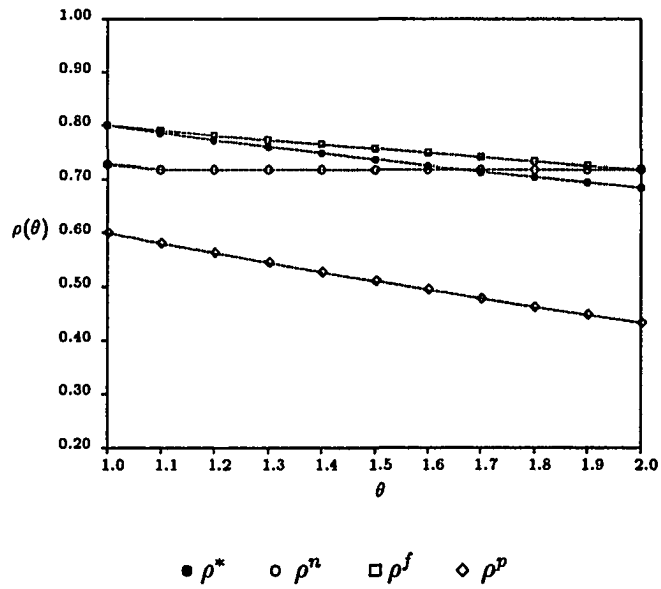


FIGURE 3.5: EXAMPLE 3.1—$\rho(\theta)$.

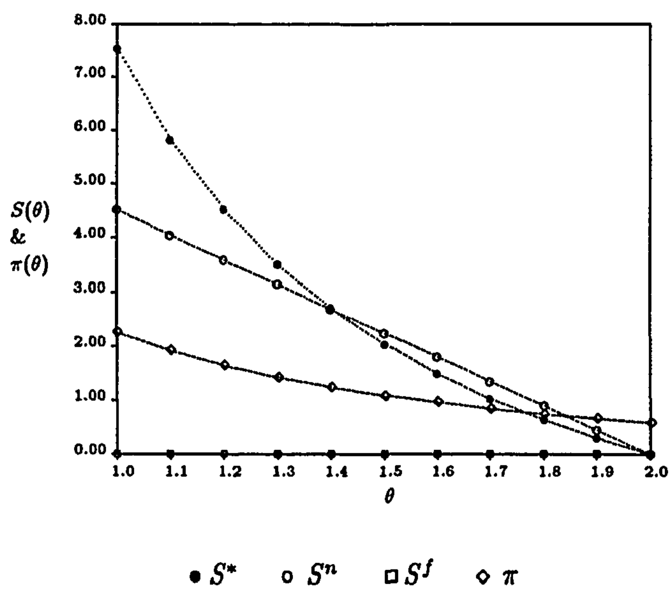


FIGURE 3.6: EXAMPLE 3.1—$S(\theta)$ & $\pi(\theta)$.

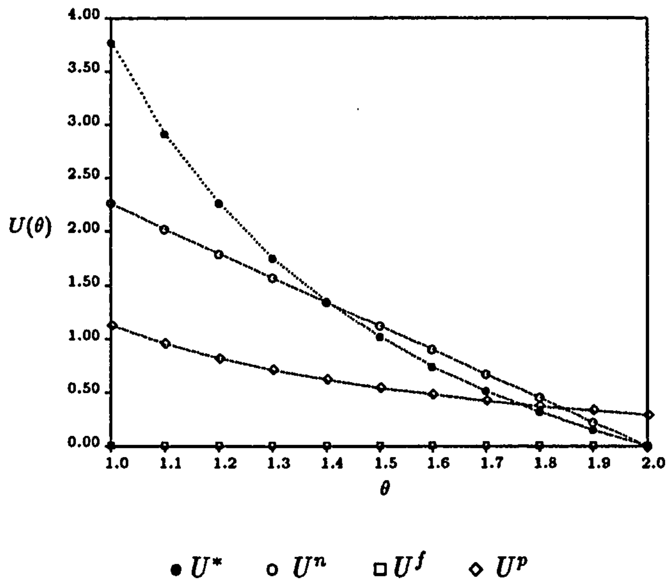


FIGURE 3.7: EXAMPLE 3.1—$U(\theta)$.

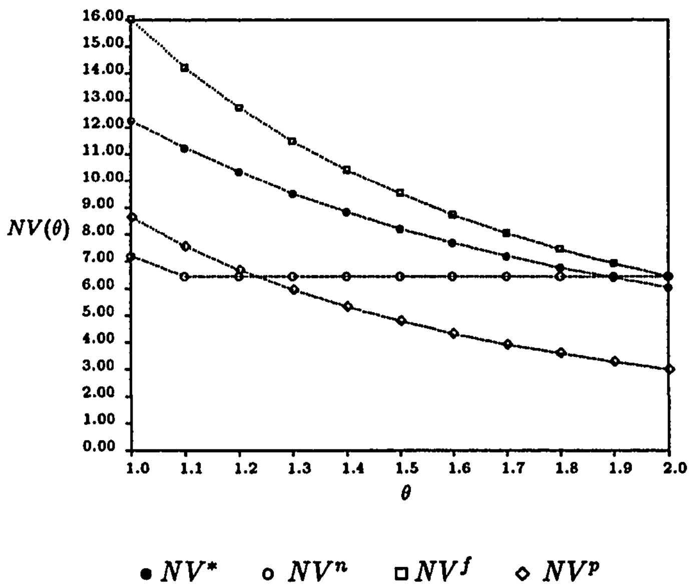


FIGURE 3.8: EXAMPLE 3.1—$NV(\theta)$.

TABLE 3.4: THE EFFECTS OF VARYING $\underline{\theta}$.

$\underline{\theta}$	2.0	1.9	1.8	1.7	1.6	1.5	1.4	1.3	1.2	1.1
$E\{NV^*\}$	6.43	6.55	6.68	6.83	6.99	7.18	7.38	7.61	7.88	8.18
$E\{NV^p\}$	3.01	3.14	3.29	3.45	3.62	3.81	4.01	4.24	4.50	4.79
$\underline{\theta}$	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
$E\{NV^*\}$	8.52	8.92	9.39	9.95	10.63	11.49	12.60	14.13	16.46	20.85
$E\{NV^p\}$	5.12	5.49	5.93	6.45	7.08	7.87	8.89	10.29	12.43	16.47

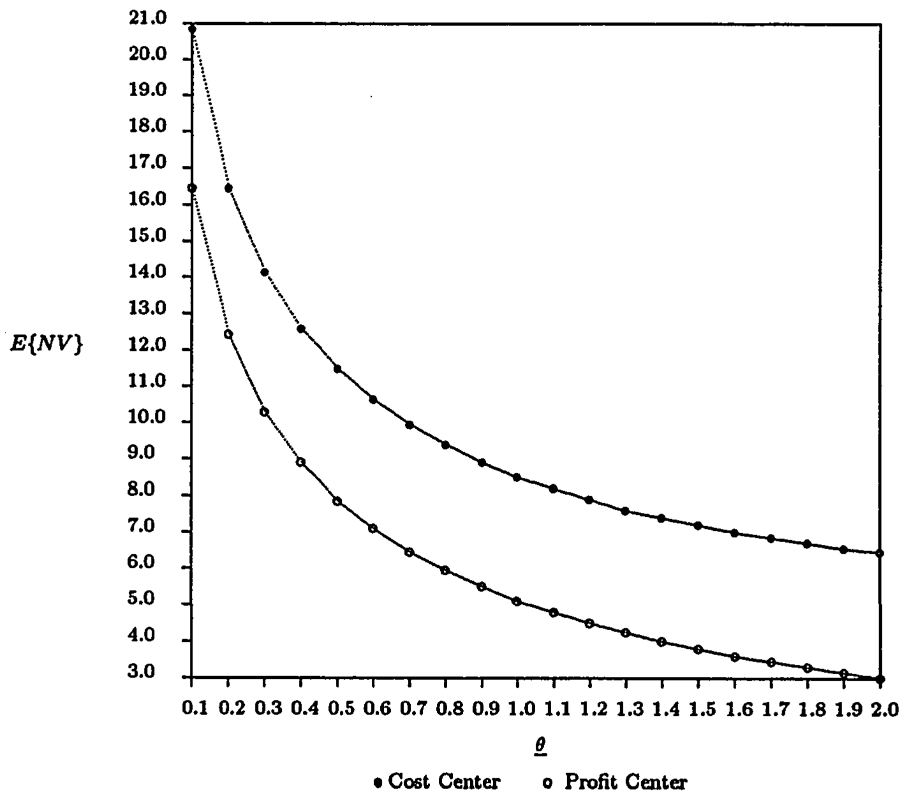


FIGURE 3.9: THE EFFECTS OF VARYING $\underline{\theta}$.

TABLE 3.5: THE EFFECTS OF VARYING $\overline{\theta}$.

$\overline{\theta}$	3.0	2.9	2.8	2.7	2.6	2.5	2.4	2.3	2.2	2.1
$E\{NV^*\}$	5.89	6.07	6.27	6.49	6.72	6.96	7.22	7.51	7.82	8.16
$E\{NV^p\}$	3.59	3.71	3.83	3.95	4.09	4.23	4.39	4.55	4.73	4.91
$\overline{\theta}$	2.0	1.9	1.8	1.7	1.6	1.5	1.4	1.3	1.2	1.1
$E\{NV^*\}$	8.52	8.93	9.37	9.87	10.42	11.04	11.75	12.57	13.52	14.64
$E\{NV^p\}$	5.12	5.34	5.57	5.83	6.12	6.43	6.77	7.16	7.59	8.07

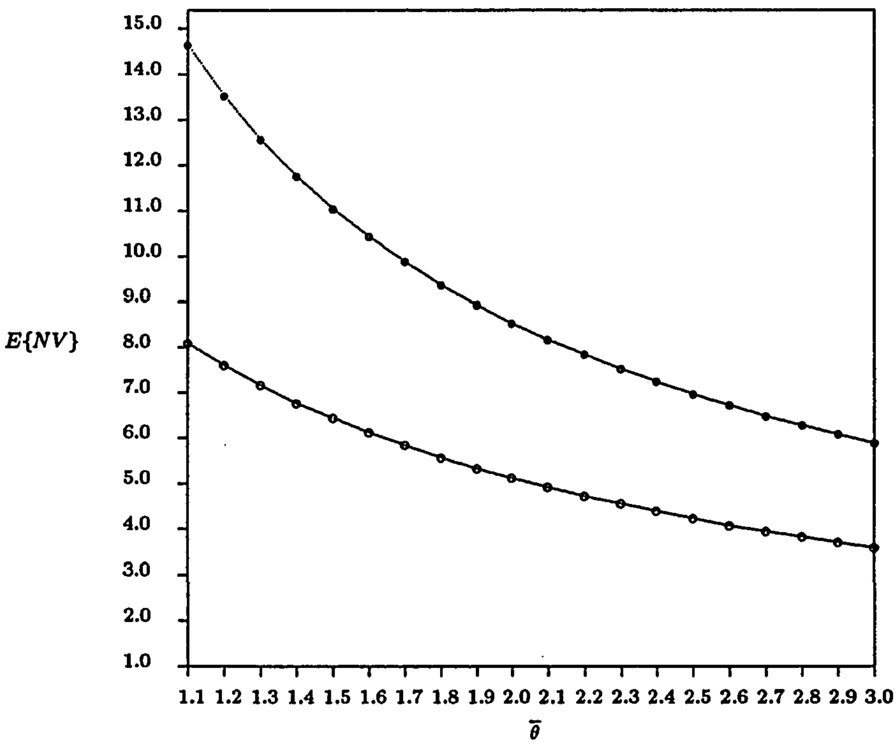


FIGURE 3.10: THE EFFECTS OF VARYING $\overline{\theta}$.

TABLE 3.6: THE EFFECTS OF MEAN-PRESERVING SPREADS.

δ	−0.4	−0.3	−0.2	−0.1	0.0	0.1	0.2
$E\{NV^*\}$	9.14	8.87	8.69	8.57	8.52	8.54	8.63
$E\{NV^p\}$	4.81	4.84	4.90	4.99	5.12	5.28	5.48
δ	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$E\{NV^*\}$	8.80	9.07	9.46	10.03	10.88	12.27	15.02
$E\{NV^p\}$	5.74	6.08	6.52	7.11	7.95	9.29	11.86

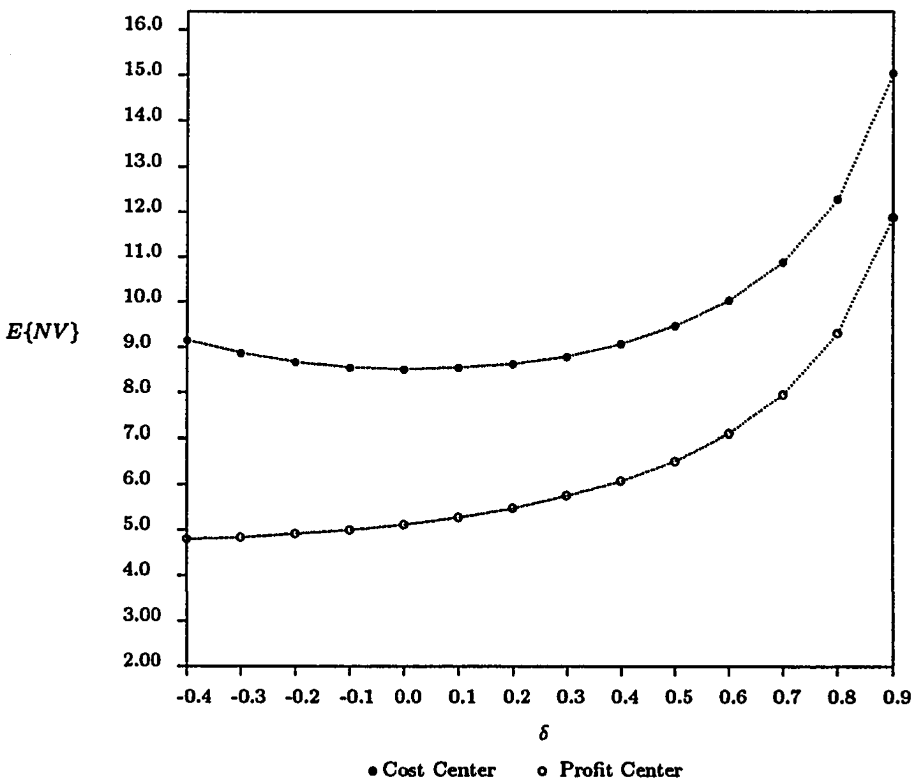


FIGURE 3.11: THE EFFECTS OF MEAN-PRESERVING SPREADS.

Table 3.7: The effects of varying ξ.

ξ	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$E\{NV^*\}$	10.04	9.70	9.37	9.07	8.79	8.52	8.27	8.04	7.82	7.61	7.41
$E\{NV^p\}$	5.71	5.59	5.47	5.36	5.24	5.12	5.00	4.88	4.76	4.64	4.52

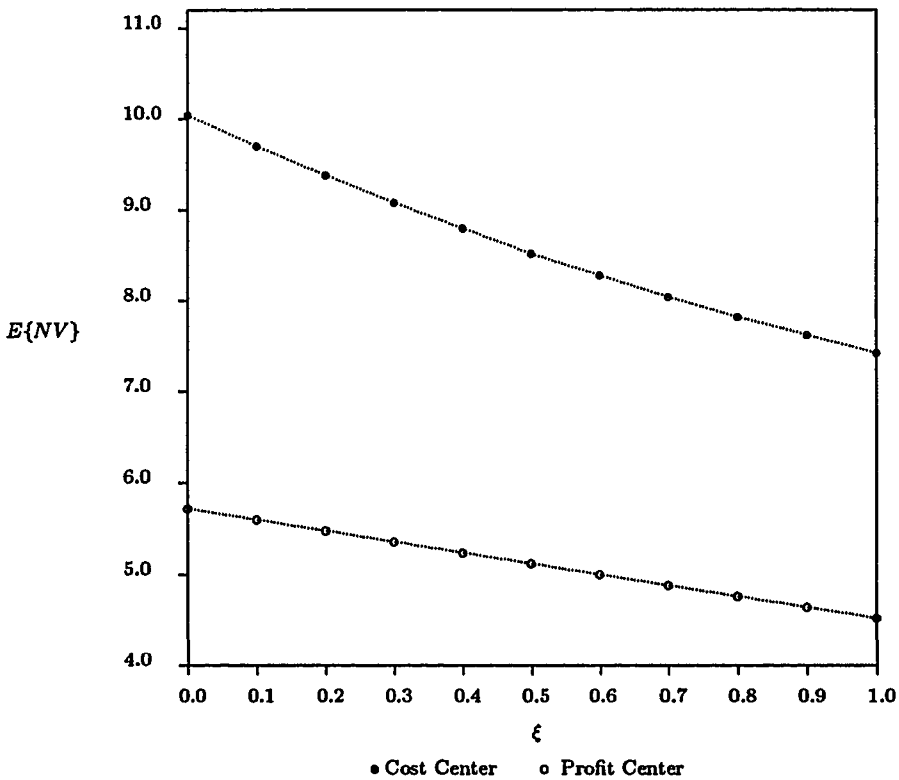


Figure 3.12: The effects of varying ξ.

TABLE 3.8: THE EFFECTS OF VARYING k.

ξ	1	2	3	4	5	6	7	8	9	10
$E\{NV^*\}$	0.00	0.34	1.85	4.57	8.52	13.69	20.08	27.69	36.53	46.59
$E\{NV^p\}$	0.00	0.00	0.67	2.46	5.12	8.64	13.03	18.28	24.41	31.40

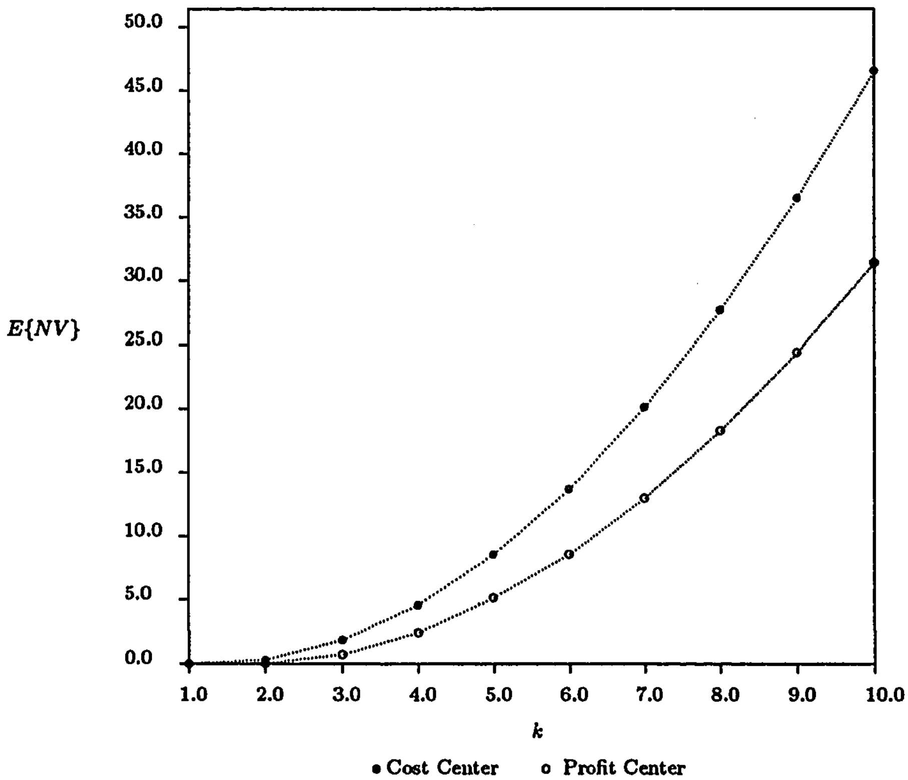


FIGURE 3.13: THE EFFECTS OF VARYING k.

Chapter 4

Finite Feasible Systems

4.1 Introduction

In the previous chapter, it is assumed that the set of feasible systems $\mathcal{K} = \mathcal{R}_+$. Consequently, under the optimal mechanism, the central management can distort the system capacity continuously with respect to the IS manager's report, $\hat{\theta}$, so as to make a continuous tradeoff between the system efficiency and the IS manager's informational rent. However, as discussed in Chapter 1, in a real-world business environment, a firm has a finite number of systems, often just a few, from which to choose. When the central management's prior beliefs about the set of possible realizations of θ is larger than $\mathcal{K}$, the central management's ability to induce the IS manager's truth-revelation is further limited, and a certain degree of pooling or bunching over Θ is inevitable. Consequently, the expected organizational net value should be reduced, and the magnitude of the reduction should depend critically on the set of feasible systems, $\mathcal{K}$. In this chapter I focus on cases where $\mathcal{K}$ is finite.

The problem associated with a finite set of feasible systems can also arise when the organization's information system can be characterized as a multi-server queueing system with a fixed capacity for each server. Under this circumstance, determining the optimal system is equivalent to determining the number of servers, and the mean delay of jobs can be written as a function of λ and the number of servers, $W(\lambda, s)$. Consequently, my analysis can be easily extended to study systems with multiple servers.

The applicability of my analysis to multi-server systems is illustrated in the example section through an $M/M/s$ queue.

Without loss of generality, I index the capacity of the set of feasible systems in decreasing order, i.e., $\mu_1 > \cdots > \mu_\kappa$ where $\kappa = |\mathcal{K}|$, the cardinality of $\mathcal{K}$. Let $K \equiv \{1, \ldots, \kappa\}$ denote the index set. I continue to assume that the IS department's cost parameter θ satisfies Assumption 1 of Chapter 2. Although $\mathcal{K}$ is discrete, I still assume that the cost for the IS department to operate an information system, $C(\mu, \theta)$, is continuously differentiable with respect to θ and satisfies the following assumption specifically for this chapter.

ASSUMPTION 4.1 1. For $\mu_1 > \mu_2$ and $\theta_1, \theta_2 \in \Theta$, if $\theta_1 > \theta_2$,

$$C(\mu_1, \theta_1) - C(\mu_2, \theta_1) > C(\mu_1, \theta_2) - C(\mu_2, \theta_2)$$

2. For all $\theta \in \Theta$, if $\mu_1 > \mu_2$, then

$$\begin{aligned} C_\theta(\mu_1, \theta) &> C_\theta(\mu_2, \theta) \\ C_{\theta\theta}(\mu_1, \theta) &\geq C_{\theta\theta}(\mu_2, \theta) \end{aligned}$$

Assumption 4.1.1 simply requires the "marginal" capacity cost to be increasing in θ. As will be seen later, if $\beta(\theta) = \frac{F(\theta)}{f(\theta)}$ is non-decreasing, Assumption 4.1 is sufficient for the "virtual" capacity cost function to be well-behaved enough so that the optimal incentive compatible mechanism can be derived by marginal analysis.

Given a system with capacity μ_i, the optimal number of jobs to serve is determined by solving the short-run problem:

$$GV(\mu_i) \stackrel{\text{def}}{=} \max_\lambda V(\lambda) - \lambda W(\lambda, \mu_i). \tag{4.1}$$

$GV(\mu_i)$ then is the *gross* organizational net value with a system of capacity μ_i before accounting for the cost of the system. The assumption that $V(\lambda)$ is concave, that $W_\lambda > 0$, and that $W_{\lambda\lambda} > 0$ implies the above problem is concave in λ, and thus it has a unique optimum for any $\mu > 0$. I further assume that $GV(\mu)$ is concave. This holds when $V(\lambda)$ is sufficiently curved. Before deriving the optimal incentive compatible mechanism, I first study the naive and full-information mechanisms, which can serve as benchmarks as in Chapter 3.

4.2 Benchmarks: The Naive and Full-information Mechanisms

The Naive Mechanism

When the central management is naive, upon receiving the IS manager's report $\hat{\theta}$, it solves the problem:

$$\max_{\mu_i : i \in K} GV(\mu_i) - C(\mu_i, \hat{\theta}). \tag{4.2}$$

Thus the central management's choice of system is a mapping: $\mu : \Theta \rightarrow \mathcal{K}$. Without loss of generality, I consider only the cases where there always exists at least one system that can yield a positive organizational net value for all $\hat{\theta}$ in Θ; i.e., there exists a $\mu_i \in \mathcal{K}$ such that $GV(\mu_i) - C(\mu_i, \hat{\theta}) > 0$ for all $\hat{\theta} \in \Theta$. Given Assumption 4.1, it is clear that the central management's capacity choice is non-increasing in $\hat{\theta}$.

Since $C(\mu, \theta)$ is continuous and monotonically increasing in θ and $\mathcal{K}$ is finite, there must be an interval, say $\Theta_t = [\theta_{t-1}, \theta_t] \subseteq \Theta$, over which, say μ_r is the solution to (4.2). Letting μ_1^n be the largest system in $\mathcal{K}$ such that there exists a $\theta \in \Theta$ that is the solution to (4.2), then any system larger than μ_1^n will never be chosen; likewise, letting μ_m^n be the smallest system in $\mathcal{K}$ such that there exists a $\theta \in \Theta$ that is the solution to (4.2), then any system smaller than μ_m^n will also never be chosen. Furthermore, for every system μ_i^n between μ_1^n and μ_m^n, there must be a $\theta_i^n \in \Theta$ such that

$$GV(\mu_i^n) - C(\mu_i^n, \theta_i^n) = GV(\mu_{i+1}^n) - C(\mu_{i+1}^n, \theta_i^n)$$

holds. That is, μ_i^n remains as the optimal system until the cost parameter becomes so large that using a smaller system can generate a larger organizational net value. When $\theta_l^n = \overline{\theta}$, because from the central management's standpoint both systems μ_l^n and μ_{l+1}^n will yield the same organizational net value, I can without loss of generality assume that the central management chooses the larger system. There must, then, be a set of adjacent systems that induces a partition that covers Θ with capacity non-increasing in $\hat{\theta}$. That is, there is a set of adjacent systems $\mathcal{K}^n \subseteq \mathcal{K}$ inducing a partition $\mathcal{P}^n \stackrel{\text{def}}{=} \{\Theta_i^n : i = 1, \ldots, |\mathcal{K}^n|\}$ such that $\Theta_1^n = [\theta_0, \theta_1^n]$ and $\Theta_i^n = (\theta_i^n, \theta_{i+1}^n]$ for all $i > 1$

where $\theta_0 = \underline{\theta}$, $\theta_{|\mathcal{K}^n|} = \overline{\theta}$, and for all $i \in \{1, \ldots, |\mathcal{K}^n| - 1\}$:

$$GV(\mu_i^n) - C(\mu_i^n, \theta_i^n) = GV(\mu_{i+1}^n) - C(\mu_{i+1}^n, \theta_i^n).$$

Let K^n denote the index set of $\mathcal{K}^n$ and

$$\mu^n(\hat{\theta}) = \arg\max_{\mu_i^n : i \in K^n} GV(\mu_i^n) - C(\mu_i^n, \hat{\theta}).$$

Since the central management is naive, the budget allocation $T^n(\hat{\theta}) \equiv C(\mu^n(\hat{\theta}), \hat{\theta})$. Then the IS manager's optimal reporting strategy under the naive mechanism is obtained by solving:

$$\max_{\hat{\theta}} C(\mu^n(\hat{\theta}), \hat{\theta}) - C(\mu^n(\hat{\theta}), \theta).$$

If the true θ falls within Θ_i^n, from the IS manager's standpoint, reporting $\hat{\theta} = \theta_i^n$ is at least as good as reporting $\hat{\theta} = \theta$ for all $\theta \in \Theta_i^n$, and so the IS manager's problem is equivalent to:

$$\max_{\theta_i : i \in K^n} C(\mu^n(\theta_i), \theta_i) - C(\mu^n(\theta_i), \theta).$$

As always, the results of a naive mechanism are intractable without specifying the functions explicitly. Further characterization of the naive mechanism is deferred to Section 4.5.

The Full-information Mechanism

Note that the optimal partition when the central management is fully informed is the same as that under the naive mechanism. Letting $\mathcal{P}^f \stackrel{\text{def}}{=} \{\Theta_i^f : i \in K^f\}$ denote the full-information optimal partition, then $K^f = K^n$ and $\mathcal{P}^f = \mathcal{P}^n$. However, under the naive mechanism, the IS manager may not have the incentive to report the cost information truthfully, and consequently the full-information solution is not implementable in general. In the next section, I derive the optimal revelation mechanism.

4.3 Revelation Mechanisms

When deriving the mechanism that maximizes the expected organizational net value, the revelation principle implies that attention can be confined to direct revelation mechanisms without loss of generality. Since $\mathcal{K}$ is finite, the central management's system decision rule necessarily maps a set of possible cost parameters into a single system. Since for each $\hat{\theta} \in \Theta$, there exists a unique λ_i that maximizes the organizational net value for every $\mu_i \in \mathcal{K}$, the central management's problem is:

$$\max_{\mu(\theta)\in\mathcal{K}} \int_{\Theta} \{GV(\mu(\theta)) - C(\mu(\theta),\theta) - \xi[T(\theta) - C(\mu(\theta),\theta)]\} \, dF(\theta)$$

subject to

$$T(\theta) - C(\mu(\theta),\theta) \geq T(\hat{\theta}) - C(\mu(\hat{\theta}),\theta), \qquad \forall\theta,\hat{\theta} \in \Theta \tag{4.3}$$

$$T(\theta) - C(\mu(\theta),\theta) \geq 0, \qquad \forall\theta \in \Theta. \tag{4.4}$$

Here, I again assume that the central management taxes away all the IS department's revenue and then allocates a lump-sum budget without loss of generality. In this subsection, I first derive the set of feasible mechanisms that satisfy (4.3) and (4.4). I then show that without loss of generality I can further focus on mechanisms that map an interval into a single system; i.e., the mechanisms induce a partition over Θ with a unique system associated with each interval.

LEMMA 4.1 *If there are two cost parameters $\theta_i, \theta_{i+1} \in \Theta$ such that $\theta_i < \theta_{i+1}$ and the central management chooses the same capacity μ_i, then the incentive compatibility constraints (4.3) imply that $\mu(\theta) = \mu_i$ and $T(\theta) = T(\theta_{i+1})$ for all $\theta \in [\theta_i, \theta_{i+1}]$.*

PROOF. Since $\mu(\theta_i) = \mu(\theta_{i+1}) = \mu_i$, $T(\theta_i) = T(\theta_{i+1})$; otherwise, $\hat{\theta}_i = \theta_{i+1}$. Suppose that there is $\theta \in (\theta_i, \theta_{i+1})$ such that the central management chooses a capacity $\mu_j \neq \mu_i$ and that the IS manager honestly reports θ. Then (4.3) implies

$$\begin{aligned} T(\theta) - C(\mu_j,\theta) &\geq T(\theta_{i+1}) - C(\mu_i,\theta) \\ T(\theta_{i+1}) - C(\mu_i,\theta_i) &\geq T(\theta) - C(\mu_j,\theta_i) \\ T(\theta_{i+1}) - C(\mu_i,\theta_{i+1}) &\geq T(\theta) - C(\mu_j,\theta_{i+1}), \end{aligned}$$

or

$$\begin{aligned} T(\theta) - T(\theta_{i+1}) &\geq C(\mu_j,\theta) - C(\mu_i,\theta) \\ T(\theta) - T(\theta_{i+1}) &\leq C(\mu_j,\theta_i) - C(\mu_i,\theta_i) \\ T(\theta) - T(\theta_{i+1}) &\leq C(\mu_j,\theta_{i+1}) - C(\mu_i,\theta_{i+1}). \end{aligned}$$

The first two inequalities imply that

$$C(\mu_i,\theta) - C(\mu_j,\theta) \geq C(\mu_i,\theta_i) - C(\mu_j,\theta_i).$$

Since $\theta > \theta_i$, the above inequality implies that $\mu_i \geq \mu_j$ by Assumption 4.1. Similarly, the first and last inequalities imply that

$$C(\mu_j,\theta_{i+1}) - C(\mu_i,\theta_{i+1}) \geq C(\mu_j,\theta) - C(\mu_i,\theta).$$

Since $\theta_{i+1} > \theta$, the above inequality implies that $\mu_j \geq \mu_i$. Therefore $\mu_j = \mu_i$ and $T(\theta) = T(\theta_{i+1})$, proving the lemma. ||

Lemma 4.1 shows that the discreteness of $\mathcal{K}$ prevents the central management from distorting the system choice continuously. Consequently, in addition to the problem of information asymmetry, the discreteness of $\mathcal{K}$ restricts the central management's ability to control the IS department even further. It is thus obvious that the expected organizational net value will be lower than in the continuous case. Furthermore, since there is no loss of generality in focusing on the mechanisms that partition Θ and assign the same capacity within each interval when seeking the direct revelation mechanisms, the "form" of the decision rule for determining the optimal system under the incentive compatible mechanisms will be the same as that under the naive mechanism.

To derive truth-revelation mechanisms, it is natural to consider the mechanisms that assign the systems to the intervals in a decreasing order, as in the naive case. I first derive the incentive compatible mechanisms for an arbitrary partition.

Let $\mathcal{P} \stackrel{\text{def}}{=} \{\Theta_i : i = 1,\ldots,\kappa\}$ be an arbitrary partition of Θ with κ intervals, where $\Theta_1 = [\underline{\theta},\theta_1]$ and $\Theta_i = (\theta_{i-1},\theta_i], i = 2,\ldots,\kappa$ such that the central management will request the IS department to acquire the system with capacity μ_i if the IS manager reports $\hat{\theta} \in \Theta_i$. That is, $|\mathcal{P}| = |\mathcal{K}| = \kappa$ and μ_i is decreasing in i.

Given an arbitrary partition and the central management's system decision rule, the central management will request the IS department to acquire a system with capacity μ_i for all $\theta \in \Theta_i$. Consequently, there is no way for the central management to distort the system choice within Θ_i in order to reduce the IS manager's informational rent, as in the continuous case. From the IS manager's standpoint, then, reporting $\hat{\theta} = \theta_i$ is at least as good as reporting $\hat{\theta} = \theta$ for all $\theta \in \Theta_i$ (since the system is the same for all $\theta \in \Theta_i$). If the central management wants to induce the IS manager's truth-revelation *within* each interval, it can do no better than allocating a budget in excess by an amount $C(\mu_i, \theta_i) - C(\mu_i, \theta)$ for each Θ_i and $\theta \in \Theta_i$; this has the same effect on the organizational net value as asking the IS manager to report $\hat{\theta} = \theta_i$ for all $\theta \in \Theta_i$ and $i \in K$. That is, inducing truth-revelation within each interval is exactly the same as asking the IS manager to always report θ_i. As a result, the upper end points of the intervals can be treated as the IS manager's message space without loss of generality. Then, given $(\mathcal{P}, \mathcal{K})$, the IS manager's reporting strategy can be recast as a mapping:

$$\sigma_1 : \Theta \rightarrow \{\theta_i : i = 1, \ldots, \kappa\},$$

and therefore without loss of generality I can again assume that the central management taxes away all the IS department's revenue. The central management's budget allocation rule is thus a mapping:

$$T : \{\theta_i : i = 1, \ldots, \kappa\} \rightarrow \mathcal{R}_+.$$

Hence, given $(T, \mathcal{P}, \mathcal{K})$, the IS manager's optimal reporting strategy is obtained by solving:

$$\max_{\theta_i : i \in K} T(\theta_i) - C(\mu_i, \theta).$$

The problem now faced by the central management is to determine the optimal way of partitioning Θ and assign one of the feasible systems to each interval.

As discussed earlier, budget allocation required to induce the IS manager's truth-revelation within each interval is the same as asking the IS manager to report the cost parameter as the upper end point of the interval, so the expected organizational net value under an incentive compatible mechanism is the same as that under a mechanism that induces the IS manager to report $\hat{\theta}(\theta) = \theta_i$, for all $\theta \in \Theta_i$ and all $i \in K$. The IS manager will report a cost parameter such that the central management will choose a system that, by design, should be assigned to the interval within which the true cost parameter lies. A mechanism $(T, \mathcal{P}, \mathcal{K})$ is thus incentive compatible if and only if

$$T_i - C(\mu_i, \theta) \geq T_j - C(\mu_j, \theta), \qquad \forall \theta \in \Theta_i, \forall i, j \in K \tag{4.5}$$

where T_i is the budget allocation when the IS manager reports θ_i. The central management's problem becomes:

$$\max_{\theta_i: i \in K} \sum_{i \in K} \int_{\theta_{i-1}}^{\theta_i} \{GV(\mu_i) - C(\mu_i,\theta) - \xi[T_i - C(\mu_i,\theta)]\} \, dF(\theta) \tag{4.6}$$

subject to

$$T_i - C(\mu_i,\theta) \geq T_j - C(\mu_j,\theta), \quad \forall \theta \in \Theta_i, \forall i,j \in K \tag{4.7}$$

$$T_i - C(\mu_i,\theta) \geq 0, \quad \forall \theta \in \Theta_i, \forall i \in K. \tag{4.8}$$

The following lemma characterizes the set of feasible mechanisms for a given pair $(\mathcal{P},\mathcal{K})$.

LEMMA 4.2 *Given $(\mathcal{P},\mathcal{K})$ with μ_i decreasing in i, a budget allocation rule is incentive compatible if and only if*

$$T_i = C(\mu_i,\theta_i) + \sum_{j=i+1}^{\kappa} [C(\mu_j,\theta_j) - C(\mu_j,\theta_{j-1})], \quad \forall i \in K, \tag{4.9}$$

where by convention

$$\sum_{j=i+1}^{\kappa} [C(\mu_j,\theta_j) - C(\mu_j,\theta_{j-1})] \equiv 0$$

when $i = \kappa$. Moreover, if

$$T_\kappa = C(\mu_\kappa,\theta_\kappa), \tag{4.10}$$

the constraint (4.8) is satisfied for all $\theta \in \Theta$.

PROOF. Given $(\mathcal{P},\mathcal{K})$, for any $i < \kappa$, local incentive compatibility requires, for all $\theta \in \Theta_i$,

$$T_i - C(\mu_i,\theta) \geq T_{i+1} - C(\mu_{i+1},\theta), \tag{4.11}$$

and for all $\theta \in \Theta_{i+1}$,

$$T_{i+1} - C(\mu_{i+1},\theta) \geq T_i - C(\mu_i,\theta). \tag{4.12}$$

Since by assumptions μ_i is decreasing in i and $C(\mu,\theta)$ satisfies Assumption 4.1, (4.11) and (4.12) imply that for all $\theta \in \Theta_i$ and for all $i \in \{1,\ldots,\kappa-1\}$,

$$T_i = T_{i+1} + C(\mu_i,\theta_i) - C(\mu_{i+1},\theta_i). \tag{4.13}$$

It now must be shown that (4.13) is also globally incentive compatible. Assuming $\theta \in \Theta_i$,

$$T_i - C(\mu_i, \theta) \geq T_j - C(\mu_j, \theta), \quad j \in \{i+1, \ldots, \kappa\}.$$

If $j = i + 1$, we are done; otherwise, let $j = i + 2$. Since from the IS manager's standpoint, claiming $\theta \in \Theta_i$ is at least as good as claiming $\theta \in \Theta_{i+1}$ by the local incentive compatibility, it is sufficient to show that for $\theta \in \Theta_i$,

$$T_{i+1} - C(\mu_{i+1}, \theta) \geq T_{i+2} - C(\mu_{i+2}, \theta). \tag{4.14}$$

From (4.13),

$$T_{i+1} = T_{i+2} + C(\mu_{i+1}, \theta_{i+1}) - C(\mu_{i+2}, \theta_{i+1}),$$

(4.14) is equivalent to requiring

$$C(\mu_{i+1}, \theta_{i+1}) - C(\mu_{i+1}, \theta) \geq C(\mu_{i+2}, \theta_{i+1}) - C(\mu_{i+2}, \theta).$$

But again by assumption that $C(\mu, \theta)$ satisfies Assumption 4.1, the above expression can be satisfied if and only if the capacity is decreasing. Following this process, it can be shown that (4.13) satisfies all the upward incentive compatibility constraints. Similarly for the downward incentive compatibility constraints, it must be shown that, for $i > 1$,

$$T_i - C(\mu_i, \theta) \geq T_j - C(\mu_j, \theta), \quad j \in \{1, \ldots, i-1\}.$$

Again if $j = i - 1$ we are done; otherwise, let $j = i - 2$, and then from (4.13),

$$T_{i-1} = T_{i-2} + C(\mu_{i-1}, \theta) - C(\mu_{i-2}, \theta) \tag{4.15}$$

is equivalent to requiring

$$C(\mu_{i-2}, \theta) - C(\mu_{i-1}, \theta) \geq C(\mu_{i-2}, \theta_{i-2}) - C(\mu_{i-1}, \theta_{i-2}).$$

Again by Assumption 4.1 the above inequality is satisfied if and only if the capacity is decreasing. Following this process, it can be shown that (4.13) also satisfies all the downward incentive compatibility constraints. Thus (4.13) satisfies all the incentive compatibility constraints.

From (4.13), it is easy to show by induction that for all $i \in \{1, \ldots, \kappa - 1\}$,

$$T_i = C(\mu_i, \theta_i) + \sum_{j=i+1}^{\kappa} [C(\mu_j, \theta_j) - C(\mu_j, \theta_{j-1})]. \tag{4.16}$$

Since the IS manager's informational rent

$$\sum_{j=i+1}^{\kappa} [C(\mu_j, \theta_j) - C(\mu_j, \theta_{j-1})]$$

is decreasing in i, setting $T_\kappa = C(\mu_\kappa, \theta_\kappa)$ will satisfy (4.8) for all $\theta \in \Theta$. ||

So in order to induce the IS manager not to overstate her department's cost, the central management needs to offset this incentive by providing the IS department some extra budget allocation:

$$C(\mu_i, \theta_i) - C(\mu_i, \theta) + \sum_{j=i+1}^{\kappa} [C(\mu_j, \theta_j) - C(\mu_j, \theta_{j-1})]$$

when $\theta \in \Theta_i$. Obviously, the above expression is the counterpart of $\int_\theta^{\bar{\theta}} C_\mu(\mu(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta}$ in Chapter 3.

The Optimal Mechanism

From Lemma 4.2, when the central management's system choice is decreasing, any mechanism $(T, \mathcal{P}, \mathcal{K})$ that satisfies (4.9) and (4.10) is feasible. Since the second term in (4.16) is independent of θ, it is straightforward to show by induction that

$$\sum_{i=1}^{\kappa} \sum_{j=i+1}^{\kappa} [C(\mu_j, \theta_j) - C(\mu_j, \theta_{j-1})] [F(\theta_i) - F(\theta_{i-1})] = \sum_{i=1}^{\kappa} [C(\mu_i, \theta_i) - C(\mu_i, \theta_{i-1})] F(\theta_{i-1}).$$

Consequently, given a feasible mechanism $(T, \mathcal{P}, \mathcal{K})$, the expected organizational net value is:

$$\sum_{i=1}^{\kappa} H_i(\mu_i, \theta_i, \theta_{i-1}) [F(\theta_i) - F(\theta_{i-1})],$$

where

$$\begin{aligned} H_i(\mu, \theta_i, \theta_{i-1}) &\stackrel{\text{def}}{=} GV(\mu_i) - VC(\mu, \theta_i, \theta_{i-1}) \\ &= GV(\mu_i) - (1 - \xi) \frac{\int_{\theta_{i-1}}^{\theta_i} C(\mu_i, \theta)\, dF(\theta)}{F(\theta_i) - F(\theta_{i-1})} \\ &\quad - \xi \left\{ C(\mu_i, \theta_i) + [C(\mu_i, \theta_i) - C(\mu_i, \theta_{i-1})] \frac{F(\theta_{i-1})}{F(\theta_i) - F(\theta_{i-1})} \right\}. \end{aligned}$$

The problem now is to derive the optimal way of partitioning Θ and determine the corresponding systems. Since it is assumed in Lemma 4.2 that the system capacity is non-increasing in θ, the decision rule which assigns the systems in non-increasing order of their capacity needs to be organizationally preferable. This decision rule is not necessarily optimal in general. In the following lemma, I show that Assumption 4.1 and the monotone hazard rate property of $\beta(\theta)$ are sufficient to guarantee that the optimal capacity assignment is non-increasing.

LEMMA 4.3 *When Assumption 4.1 holds and $\beta(\theta)$ is non-decreasing, the optimal capacity decision is non-increasing in i.*

PROOF. For any fixed $\mu > 0$, the virtual capacity cost for interval i:

$$\begin{aligned}
VC(\mu,\theta_i,\theta_{i-1}) &\stackrel{\text{def}}{=} (1-\xi)\frac{\int_{\theta_{i-1}}^{\theta_i} C(\mu,\theta)\,dF(\theta)}{F(\theta_{i-1})-F(\theta_{i-1})} \\
&\quad +\xi\left\{C(\mu,\theta_i)+[C(\mu,\theta_i)-C(\mu,\theta_{i-1})]\frac{F(\theta_{i-1})}{F(\theta_i)-F(\theta_{i-1})}\right\} \\
&= \frac{(1-\xi)\int_{\theta_{i-1}}^{\theta_i} C(\mu,\theta)\,dF(\theta)+\xi[F(\theta_i)C(\mu,\theta_i)-F(\theta_{i-1})C(\mu,\theta_{i-1})]}{F(\theta_i)-F(\theta_{i-1})} \\
&= \frac{\int_{\theta_{i-1}}^{\theta_i}\{(1-\xi)C(\mu,\theta)+\xi\,[C(\mu,\theta)+\beta(\theta)C_\theta(\mu,\theta)]\}\,dF(\theta)}{\int_{\theta_{i-1}}^{\theta_i} dF(\theta)} \\
&= \frac{\int_{\theta_{i-1}}^{\theta_i}\{C(\mu,\theta)+\xi\beta(\theta)C_\theta(\mu,\theta)\}\,dF(\theta)}{\int_{\theta_{i-1}}^{\theta_i} dF(\theta)}. \qquad (4.17)
\end{aligned}$$

By Assumption 4.1 and the hypothesis that $\beta(\theta)$ is non-decreasing, for an arbitrary partition,

$$C_\theta(\mu,\theta)+\xi[\beta'(\theta)C_\theta(\mu,\theta)+\beta(\theta)C_{\theta\theta}(\mu,\theta)]$$

is increasing in μ, and thereby the optimal capacity assignment is non-increasing in i. ||

Lemma 4.3 shows that the condition required for the system decision rule to be globally incentive compatible is the same as that required in the continuous case. Furthermore, from (4.17), the effect of the discreteness of $\mathcal{K}$ is apparent. When $\mathcal{K}$ is continuous, the central management can, for each θ, prescribe an appropriate system, and thereby μ in (4.17) is a continuous function of θ. But for the finite case, the system is fixed over each interval. As a result, the central management loses the ability to make a *continuous* tradeoff between the IS manager's informational rent and the operational efficiency by distorting the capacity continuously. Given a mechanism $(\mathcal{T},\mathcal{P},\mathcal{K})$ in which μ_i is decreasing, (4.9) and (4.10) will induce the IS manager not to lie across intervals. Then, for a given $\mathcal{K}$, the remaining task for the central management is to find the optimal way of partitioning Θ and determine the corresponding systems. The central management's problem now becomes:

$$\max_{\theta_i : i \in K} \sum_{i \in K} H(\mu_i, \theta_i, \theta_{i-1}) \left[F(\theta_i) - F(\theta_{i-1}) \right]. \tag{4.18}$$

I now derive the optimal mechanism. As in the naive case, because $VC(\mu,\theta_i,\theta_{i-1})$ is continuous in both θ_i and θ_{i-1}, it is clear that, at optimum, the systems having positive probability of being chosen must consist of a set of adjacent systems. The following proposition characterizes the optimal incentive compatible mechanism.

PROPOSITION 4.1 *Suppose Assumption 4.1 holds and $\beta(\theta)$ is non-decreasing. Letting $\{\theta_1^*,\ldots,\theta_\kappa^*\}$ denote the optimal partition, then θ_i^* can be determined independent of $\theta_j^*, j \neq i$, and is given by the solution to the following equation:*

$$\begin{aligned} 0 &= GV(\mu_i) - GV(\mu_{i+1}) - C(\mu_i,\theta_i) + C(\mu_{i+1},\theta_i) \\ &\quad -\xi\beta(\theta_i)\left[C_\theta(\mu_i,\theta_i) - C_\theta(\mu_{i+1},\theta_i)\right], \end{aligned} \tag{4.19}$$

provided that it yields a solution in Θ.

PROOF. Lemma 4.3 shows that the optimal capacity assignment is decreasing in i, and consequently θ_i^* must be increasing. First note that (see Lemma 4.3)

$$\begin{aligned} \frac{\partial VC(\mu,\theta_i,\theta_{i-1})}{\partial \theta_i} &= \frac{f(\theta_i)}{\int_{\Theta_i} dF(\theta)}\left[C(\mu,\theta_i) + \xi\beta(\theta_i)C_\theta(\mu,\theta_i) - VC(\mu,\theta_i,\theta_{i-1})\right] > 0; \\ \frac{\partial VC(\mu,\theta_i,\theta_{i-1})}{\partial \theta_{i-1}} &= -\frac{f(\theta_{i-1})}{\int_{\Theta_i} dF(\theta)}\left[C(\mu,\theta_{i-1}) + \xi\beta(\theta_{i-1})C_\theta(\mu,\theta_{i-1}) - VC(\mu,\theta_i,\theta_{i-1})\right] > 0, \end{aligned}$$

so the virtual system cost is increasing in both θ_i and θ_{i+1}. To derive the optimal mechanism, differentiating (4.18) with respect to θ_i for all i and equating the expressions to zero yields the following first-order conditions:

$$0 = \left[H(\mu_i,\theta_i,\theta_{i-1}) - H(\mu_{i+1},\theta_{i+1},\theta_i)\right] f(\theta_i) + \frac{\partial H(\mu_i,\theta_i,\theta_{i-1})}{d\theta_i}\left[F(\theta_i) - F(\theta_{i-1})\right]$$

$$
\begin{aligned}
&+\frac{\partial H(\mu_{i+1},\theta_{i+1},\theta_i)}{d\theta_i}[F(\theta_{i+1})-F(\theta_i)] \\
= \; & [H(\mu_i,\theta_i,\theta_{i-1})-H(\mu_{i+1},\theta_{i+1},\theta_i)]f(\theta_i)-\frac{\partial VC(\mu_i,\theta_i,\theta_{i-1})}{\partial\theta_i}[F(\theta_i)-F(\theta_{i-1})] \\
&-\frac{\partial VC(\theta_{i+1},\theta_i)}{d\theta_i}[F(\theta_{i+1})-F(\theta_i)] \qquad (4.20) \\
= \; & f(\theta_i)\{GV(\mu_i)-VC(\mu_i,\theta_i,\theta_{i-1})-GV(\mu_{i+1})+VC(\mu_{i+1},\theta_{i+1},\theta_i) \\
&-[C(\mu_i,\theta_i)+\xi\beta(\theta_i)C_\theta(\mu_i,\theta_i)-VC(\mu_i,\theta_i,\theta_{i-1})] \\
&\quad+[C(\mu_{i+1},\theta_i)+\xi\beta(\theta_i)C_\theta(\mu_{i+1},\theta_i)-VC(\mu_i,\theta_{i+1},\theta_i)]\} \\
= \; & f(\theta_i)\{GV(\mu_i)-GV(\mu_{i+1})-C(\mu_i,\theta_i)+C(\mu_{i+1},\theta_i) \\
&-\xi\beta(\theta_i)[C_\theta(\mu_i,\theta_i)-C_\theta(\mu_{i+1},\theta_i)]\}. \qquad (4.21)
\end{aligned}
$$

Since $f(\theta) > 0$ for all $\theta \in \Theta$, all the critical points are the solutions to the equation (4.19), which is a function of θ_i alone. It is easy to verify that the right-hand side of (4.19) is monotonically decreasing in θ_i provided that Assumption 4.1 holds and $\beta(\theta)$ is non-decreasing, and thereby (4.19) yields a unique solution. ||

The last two terms in (4.20) are the effects on the expected informational rent of the IS manager on Θ_i and Θ_{i+1}, respectively, as θ_i increases. In contrast to the naive case, the optimal mechanism must account for the effects of varying θ_i's on the IS manager's informational rent. Since both $\frac{\partial VC(\mu_i,\theta_i,\theta_{i-1})}{\partial\theta_i}$ and $\frac{\partial VC(\mu_{i+1},\theta_{i+1},\theta_i)}{\partial\theta_i}$ are positive, an increase in θ_i increases the expected informational rent if θ falls within either Θ_i or Θ_{i+1}. Consequently, when determining the optimal partition, the central management must consider changes in the (virtual) organizational net value $H(\mu_i,\theta_i,\theta_{i-1}) - H(\mu_{i+1},\theta_{i+1},\theta_i)$ as well as the expected IS manager's informational rent. (4.20) shows, at optimum, $H(\mu_i,\theta_i,\theta_{i-1})$ must be large enough to balance out both of the effects that reduce the expected organizational net value. However, at optimum, the terms involving the upper and lower end points cancel out each other. Since μ_i's are given and cannot be distorted as in the continuous case, $GV(\mu_i)$ is a constant over Θ_i. As a result, the set of first-order conditions characterizing the optimal partition is not linked, and each first-order condition can be solved independently provided that it yields a solution in Θ.

Moreover, since $f(\theta) > 0$ for all θ, (4.21) is equivalent to:

$$GV(\mu_i) - GV(\mu_{i+1}) = C(\mu_i, \theta_i) - C(\mu_{i+1}, \theta_i) + \xi\beta(\theta_i)\left[C_\theta(\mu_i, \theta_i) - C_\theta(\mu_{i+1}, \theta_i)\right]. \quad (4.22)$$

So, given μ_i and μ_{i+1}, the optimal θ_i is decreasing in ξ provided that Assumption 4.1 holds and that $\beta(\theta)$ is strictly increasing. Obviously, there may well be some systems in $\mathcal{K}$ such that (4.22) cannot hold as an identity. This can happen when a system is either too large or too small. When this occurs, all such systems at the upper and lower ends will be excluded.

Let μ_1^* and μ_2^* be the two largest systems in $\mathcal{K}$ with a realization $\theta_1^* \in \Theta$ such that the equation:

$$GV(\mu_1^*) - GV(\mu_2^*) = C(\mu_1^*, \theta_1^*) - C(\mu_2^*, \theta_1^*) + \xi\beta(\theta_1)\left[C_\theta(\mu_1^*, \theta_1^*) - C_\theta(\mu_2^*, \theta_1^*)\right]$$

can hold. Then the optimal partition induced by the optimal mechanism can be obtained by repeatedly applying (4.22) with smaller capacities until either all feasible systems are exhausted or there exist no smaller systems such that (4.22) can hold. Let $\mathcal{P}^* \stackrel{\text{def}}{=} \{\Theta_i^* : i \in K^*\}$ be the optimal partition induced by the optimal mechanism and $\mathcal{K}^*$ be the corresponding optimal set of systems, then the expected organizational net value under the optimal mechanism is:

$$\sum_{i \in K^*} \int_{\Theta_i^*} H(\mu_i^*, \theta_i^*, \theta_{i-1}^*)\, dF(\theta). \quad (4.23)$$

The effect of the IS manager's informational rent can now be clearly seen. Compared with the full-information case, the IS manager's informational rent affects the capacity decision through the extra nonnegative term:

$$\xi\beta(\theta_i)\left[C_\theta(\mu_i^*, \theta_i) - C_\theta(\mu_{i+1}^*, \theta_i)\right]$$

in (4.22), which equals 0 if $\xi = 0$, $\theta_i = \underline{\theta}$, or both. Thus, when $\xi > 0$, $\theta_i^f < \theta_i^*$ for all $i \in K^f$. In other words, the presence of information asymmetry forces the central management to distort the partition downward in an attempt to reduce the IS manager's informational rent. Consequently, some systems with too small a capacity to be included in the full-information solution may be included in the optimal mechanism when the information is incomplete.

The following proposition characterizes the optimal partition for the special case where $C(\mu,\theta)$ is multiplicatively separable in the cost parameter and capacity. For this special case, the first-order conditions (4.20) can be reduced to a very simple form in which the optimal θ_i^*'s can be obtained by inverting a monotonic function.

PROPOSITION 4.2 *Suppose that the capacity function is multiplicatively separable taking the form: $C(\mu,\theta)=\tau(\theta)c(\mu)$, where both $\tau(\theta)$ and $c(\mu)$ are increasing and (weakly) convex, and $\beta(\theta)$ is non-decreasing, then*

$$\theta_i^*(A(\mu_i,\mu_{i+1}))\equiv h^{-1}(A(\mu_i,\mu_{i+1})), \tag{4.24}$$

provided $\theta_i^*(A(\mu_i,\mu_{i+1}))\in\Theta$, where

$$\begin{aligned} h(\theta) &= \tau(\theta)+\xi\tau'(\theta)\beta(\theta) \\ A(\mu_i,\mu_{i+1}) &= \frac{GV(\mu_i)-GV(\mu_{i+1})}{c(\mu_i)-c(\mu_{i+1})}. \end{aligned}$$

PROOF. Since the assumptions of this proposition are a special case of Assumption 4.1, Proposition 4.1 applies and only (4.24) must be shown to hold. Let

$$\gamma(\theta_i,\theta_{i-1};\xi)\stackrel{\text{def}}{=}\frac{\int_{\theta_{i-1}}^{\theta_i}\{\tau(\theta)+\xi\beta(\theta)\tau'(\theta)\}\,dF(\theta)}{\int_{\theta_{i-1}}^{\theta_i}dF(\theta)},$$

then given the hypotheses, (4.20) reduces to

$$\begin{aligned} 0 \;=\; & \{GV(\mu_i)-\gamma(\theta_i,\theta_{i-1};\xi)c(\mu_i)-[GV(\mu_{i+1})-\gamma(\theta_{i+1},\theta_i;k)c(\mu_{i+1})]\}\,f(\theta_i) \\ & -\{[\tau(\theta_i)+\xi\tau'(\theta_i)\beta(\theta_i)]\,f(\theta_i)-\gamma(\theta_i,\theta_{i-1};\xi)f(\theta_i)\}\,c(\mu_i) \\ & +\{[\tau(\theta_i)+\xi\tau'(\theta_i)\beta(\theta)]\,f(\theta_i)-\gamma(\theta_{i+1},\theta_i;\xi)f(\theta_i)\}\,c(\mu_{i+1}), \end{aligned}$$

which is equivalent to:

$$0=GV(\mu_i)-GV(\mu_{i+1})-[\tau(\theta_i)+\xi\tau'(\theta_i)\beta(\theta)]\,[c(\mu_i)-c(\mu_{i+1})],$$

or

$$h(\theta_i)=\tau(\theta_i)+\xi\tau'(\theta_i)\beta(\theta_i)=\frac{GV(\mu_i)-GV(\mu_{i+1})}{c(\mu_i)-c(\mu_{i+1})}. \tag{4.25}$$

Since the left-hand side of (4.25) is strictly increasing and by assumption $GV(\cdot)$ is concave, the right-hand side of (4.25) is strictly decreasing, so θ_i^* can be uniquely determined by inverting (4.25), provided that $h^{-1}(A(\mu_i, \mu_{i+1})) \in \Theta$. ||

For example, if $\tau(\theta) = \theta$ and θ is distributed uniformly over Θ, then from Proposition 4.2,

$$\theta_i^* = \frac{1}{1+\xi}\left[\frac{GV(\mu_i) - GV(\mu_{i+1})}{c(\mu_i) - c(\mu_{i+1})} + \xi\underline{\theta}\right],$$

provided that $\theta_i^* \in \Theta$. For the full-information case,

$$\theta_i^f = \frac{GV(\mu_i) - GV(\mu_{i+1})}{c(\mu_i) - c(\mu_{i+1})}.$$

Clearly, given $\theta_i^f \in (\underline{\theta}, \overline{\theta}]$, $\theta_i^* < \theta_i^f$.

4.4 Profit Center

When the IS department is organized as a profit center, for a given μ_i, let $R(\mu_i)$ denote the value of the following short-run problem:

$$R(\mu_i) \stackrel{\text{def}}{=} \max_{\lambda} \lambda(V'(\lambda) - W(\lambda, \mu_i)). \tag{4.26}$$

Then for a realized θ, the IS department's profit maximizing capacity choice is the solution to the following problem:

$$\pi(\theta) = \max_{\mu_i : i \in K} R(\mu_i) - C(\mu_i, \theta).$$

Let

$$\mu^p(\theta) = \arg\max_{\mu_i : i \in K} R(\mu_i) - C(\mu_i, \theta)$$

and

$$\lambda^p(\theta) \equiv \lambda(\mu^p(\theta)) = \arg\max_{\lambda} \lambda(V'(\lambda) - W(\lambda, \mu^p(\theta))).$$

Let θ_1^p be the smallest realization in Θ with two systems $\mu_j, \mu_{j+1} \in \mathcal{K}$ such that the equation:

$$R(\mu_j) - C(\mu_j, \theta_1^p) = R(\mu_{j+1}) - C(\mu_{j+1}, \theta_1^p)$$

can hold. Then the optimal partition for the profit center can be obtained by repeatedly applying the above equation as for the optimal mechanism. Let $\mathcal{P}^p \stackrel{\text{def}}{=} \{\Theta_i^p : i \in K^p\}$ denote the profit center's profit maximizing partition; the expected organizational net value when the IS department is organized as a profit center is:

$$\sum_{i \in K^p} \int_{\Theta_i^p} NV^p(\theta)\, dF(\theta),$$

where

$$\begin{aligned} NV^p(\theta) &= V(\lambda^p(\theta)) - \lambda^p(\theta) W(\lambda^p(\theta), \mu^p(\theta)) - C(\mu^p(\theta), \theta) - \xi\pi(\theta) \\ &= V(\lambda^p(\theta)) - \lambda^p(\theta) V'(\lambda^p(\theta)) + (1-\xi)\pi(\theta). \end{aligned}$$

Clearly, when $\xi = 1$, the above expression reduces to:

$$NV^p(\theta) = V(\lambda^p(\theta)) - \lambda^p(\theta) V'(\lambda^p(\theta)),$$

which is positive provided that $V(\lambda)$ is strictly concave.

4.5 Examples

I make use of Example 3.1 in Chapter 3 to evaluate the expected organizational net value that can be generated under each mechanism for a single-server queue and a multi-server queue.

Example 4.1: A Single Server Queue—$M/M/1$

In this example, I further make the following specific assumptions:

1. $V(\lambda) = 2k\sqrt{\lambda}$ with $k = 5$.
2. $C(\mu, \theta) = \theta\mu$.
3. θ is distributed uniformly over [1, 2].
4. $E\{D(\tilde{W}(\lambda, \mu))\} = E\{\tilde{W}(\lambda, \mu)\} = (\mu - \lambda)^{-1}$ and $\xi = 1$.
5. $\mathcal{K} = \{30, 8, 6, 2\}$.

TABLE 4.1: EXAMPLE 4.1—THE GROSS ORGANIZATIONAL NET VALUE AND THE OPTIMAL ARRIVAL RATE.

μ	30	8	6	2
$GV(\mu)$	45.04337	21.49529	18.22642	9.547974
$\lambda(\mu)$	24.54772	6.01876	4.41234	1.32185

TABLE 4.2: EXAMPLE 4.1—THE EXCESS BUDGET ALLOCATION THAT THE IS MANAGER CAN OBTAIN OVER EACH INTERVAL UNDER THE NAIVE MECHANISM.

Θ_i^n	$S^n(\theta_1^n;\theta)$	$S^n(\theta_2^n;\theta)$	$S^n(\theta_3^n;\theta)$
$[\theta_0^n,\theta_1^n] = [1.00000, 1.07368]$	[2.21040, 0.00000]	[5.07544, 4.48600]	[6.00000, 5.55792]
$(\theta_1^n,\theta_2^n] = (1.07368, 1.63443]$	—	(4.48600, 0.00000]	(5.55792, 2.19342]
$(\theta_2^n,\theta_3^n] = (1.63443, 2.00000]$	—	—	(2.19342, 0.00000]

I first calculate the $GV(\mu_i)$'s. Since (4.1) is concave, for a given capacity μ_i, the solution to the first-order condition of (4.1):

$$0 = \frac{5}{\sqrt{\lambda}} - \frac{\mu_i}{(\mu_i - \lambda)^2}$$

yields a unique global maximum. The corresponding $GV(\mu_i)$'s and $\lambda(\mu_i)$'s are given in Table 4.1.

The Naive Mechanism. By setting

$$GV(\mu_i) - \theta_i\mu_i = GV(\mu_{i+1}) - \theta_i\mu_{i+1}, \tag{4.27}$$

θ_i^n's can be calculated as shown in the first column of Table 4.2. Note that, from (4.27), $\theta_3^n = 2.16961 > 2$, so the smallest system μ_4 will be excluded and $\theta_3^n = 2$. Of course, the resulting partition under the naive mechanism is the same as that of the full-information case.

Given the partition, the IS manager must decide what to report for a given θ. But there are only three points that the manager needs to consider, namely, θ_1^n, θ_2^n, and $\theta_3^n (= \overline{\theta})$. Table 4.2 also shows the ranges of the informational rent that the IS manager can command by misrepresentation. For example, depending on θ, the IS manager can obtain an informational rent over the range [6.00000, 5.55792] for $\theta \in [\theta_0^n, \theta_1^n]$ if she reports $\hat{\theta} = \theta_3^n = \overline{\theta}$. From Table 4.2, it is easy to verify that $\hat{\theta} = \theta_3^n$ for all $\theta \in \Theta$. As a result, the central management will always request the IS department to acquire $\mu_3 (= 6)$, and then for all $\theta \in \Theta$, the ex post organizational net value, and thereby the expected organizational net value, is:

$$\begin{aligned} NV(\mu_3) &= GV(\mu_3) - \overline{\theta}\mu_3 \\ &= 18.22642 - 2 \times 6 \\ &= 6.22642, \end{aligned}$$

and the IS manager's information rent is:

$$S^n(\overline{\theta}; \theta) = (2 - \theta)6.$$

Notice that this example is the same as Example 3.1 in Chapter 3. In comparison, the expected organizational net value decreases from about 6.45 in the continuous case to 6.23 in the current finite case. Although the magnitude of the reduction in the net value is not very significant for this particular example, it could be substantial for another $\mathcal{K}$. Also the resulting net value for this example is exactly the same as when $\mathcal{K} = \{6\}$. Further notice that a larger $\mathcal{K}$ does not necessarily generate a larger organizational net value under the naive mechanism unless the smaller set of systems is a subset of the larger one. That is, for any two sets of systems $\mathcal{K}_1$ and $\mathcal{K}_2$ such that $\mathcal{K}_1 \subset \mathcal{K}_2$, the expected organizational net value that which can be generated by $\mathcal{K}_2$ must be at least as large as that which can be generated by $\mathcal{K}_1$. For instance, when there is only one feasible system with capacity 5, the IS manager will still always report the cost parameter as 2, and it is easy to verify that the organizational net value is:

$$16.40311 - 2 \times 5 = 6.40311.$$

Of course, if there are some additional feasible systems, the expected organizational net value will be at least as large as when there is only one feasible system with capacity

TABLE 4.3: EXAMPLE 4.1—THE TOTAL BUDGET ALLOCATION TO THE IS DEPARTMENT UNDER THE OPTIMAL MECHANISM.

i	1	2	3	4
θ_i^*	1.03518	1.31722	1.58481	2.00000
i	1	2	3	4
T_i^*	35.74770	12.97366	10.33923	4.00000

5. On the other hand, when a system with capacity 30 is the only feasible system, the organization will incur a negative organizational net value with this system whenever the realization of θ is greater than 1.50146. Even when the realization is less than 1.50146, the organization can only obtain a zero organizational net value, since the IS manager will always report $\hat{\theta} = 1.50146$.

The Optimal Mechanism. By Proposition 4.2, the optimal mechanism is determined by solving the following equations:

$$0 = GV(\mu_i) - GV(\mu_{i+1}) - (2\theta_i - \underline{\theta})(\mu_i - \mu_{i+1}). \tag{4.28}$$

Rearranging the terms gives:

$$\theta_i^* = \frac{\underline{\theta}}{2} + \frac{GV(\mu_i) - GV(\mu_{i+1})}{2(\mu_i - \mu_{i+1})}. \tag{4.29}$$

To check the second-order conditions, differentiating the right-hand side of (4.28) with respect to θ_i yields:

$$-2(\mu_i - \mu_{i+1}) < 0,$$

since $\mu_i > \mu_{i+1}$. So (4.18) is concave in the θ_i's, and thereby the globally optimal solution is given by (4.29) as long as the resulting θ_i's are in Θ. Hence

$$\frac{GV(\mu_i) - GV(\mu_{i+1})}{\mu_i - \mu_{i+1}}$$

is increasing in i because μ_i is decreasing in i, and thus θ_i^* is decreasing in system capacity.

Given $\{\mathcal{K}\} = \{30, 8, 6, 2\}$, $\Theta = [1, 2]$, and

$$T_i^* = \theta_i^* \mu_i^* + \sum_{j=i+1}^{4} (\theta_j^* - \theta_{j-1}^*)\mu_j^*,$$

the optimal partition and the resulting optimal incentive compatible budget allocation rule are as shown in Table 4.3. Thus, for each $\theta \in \Theta_i^*$, the IS manager's informational rent is:

$$\begin{aligned} S_i^*(\theta) &= T_i^* - \theta\mu_i^* \\ &= (\theta_i^* - \theta)\mu_i^* + \sum_{j=i+1}^{4} (\theta_j^* - \theta_{j-1}^*)\mu_j^*. \end{aligned}$$

Direct calculation shows that, under the optimal mechanism, the expected organizational net value:

$$\begin{aligned} E\{NV^*(\theta)\} &= \sum_{i \in K^*} \int_{\Theta_i^*} NV^*(\theta)\, dF(\theta) \\ &= \sum_{i \in K^*} [GV(\mu_i^*) - (\theta_i^* + \theta_{i-1}^* - \underline{\theta})\mu_i^*][F(\theta_i^*) - F(\theta_{i-1}^*)] \\ &= 7.14445, \end{aligned}$$

which is much larger than under the naive mechanism, as expected. When $\mathcal{K}$ is continuous as in Chapter 3, the expected organizational net value equals 7.41215. Hence, for this particular example, the discreteness of $\mathcal{K}$ does not have a very significant impact on the expected organizational net value.

The Profit Center. For the profit center case, I first calculate the IS department's maximum revenue for a given system by solving:

$$R(\mu_i) = \max_{\lambda} k\sqrt{\lambda} - \frac{\lambda}{\mu_i - \lambda},$$

which yields the first-order condition:

$$0 = \frac{5}{\sqrt{\lambda}} - \frac{\mu_i}{(\mu_i - \lambda)^2}.$$

Then numerically

$$R(30) = 20.71721$$

TABLE 4.4: EXAMPLE 4.1—THE PARTITIONS OF THE FOUR CASES CONSIDERED.

	θ_0	θ_1	θ_2	θ_3	θ_4
Naive Mechanism	1.00000	1.07368	1.63443	2.00000	—
Optimal Mechanism	1.00000	1.03518	1.31722	1.58481	2.00000
Full-Information	1.00000	1.07368	1.63443	2.00000	—
Profit Center	1.00000	2.00000	—	—	—

$$
\begin{aligned}
R(8) &= 9.54797 \\
R(6) &= 8.02022 \\
R(2) &= 4.02238.
\end{aligned}
$$

It is easy to verify that, for all $\theta \in \Theta$, the profit maximizing capacity is $\mu_1^p = \mu_4(= 2)$ with $\lambda^p(2) = 1.08677$, which always yields a positive profit for the IS department. Since $\xi = 1$ and $V(\lambda)$ is homogeneous of degree $\frac{1}{2}$, the expected organizational net value is:

$$
\begin{aligned}
E\{NV^p(\theta)\} &= \int_\Theta V(\lambda^p(2)) - \lambda^p(2)V'(\lambda^p(2))dF(\theta)) \\
&= \frac{1}{2}V(\lambda^p(2)) \\
&= k\sqrt{\lambda^p(2)} \\
&= 5.21242,
\end{aligned}
$$

which is smaller than it is under either the naive or the optimal mechanism.

Discussion. Table 4.4 summarizes the partitions under the four mechanisms considered. Compared with the full-information case, we see that θ_i^*'s are shifted downward and all four systems have positive probability of being chosen *ex ante*. The presence of information asymmetry thus forces the central management to distort intervals downward in order to reduce the IS manager's informational rent. Furthermore, the system (with capacity 2) excluded from the full-information mechanism is included in the optimal mechanism. That is, after accounting for the IS manager's informational rent, the system that is originally inferior regardless of the realization of θ becomes organizationally preferable when the cost parameter is sufficiently large. Again, because

TABLE 4.5: EXAMPLE 4.1—THE RESULTING EX POST CAPACITY AND THE EXPECTED ORGANIZATIONAL NET VALUE.

Interval	Θ_1	Θ_2	Θ_3	Θ_4	$E\{NV(\theta)\}$
Naive Mechanism	6	6	6	—	6.22642
Optimal Mechanism	30	8	6	2	7.14445
Fully Informed Case	30	8	6	—	9.68327
Profit Center	2	—	—	—	5.21242

the additional informational rent required for inducing truth-revelation outweighs the improvement in operational efficiency, the full-information partition is feasible but not optimal. In fact, it is easy to verify that the expected organizational net value equals 6.22103 when the full-information partition is truthfully implemented.

An interesting phenomenon to observe is that the standard "no distorting at the bottom" property in the continuous case does not hold in general in the discrete case. Figure 4.1 shows that the system capacity is not distorted over both the intervals $[\underline{\theta}, \theta_1^*]$ and $(\theta_1^f, \theta_2^*]$. This illustrates how the discreteness of $\mathcal{K}$ constrains the central management's ability to control the IS department. Nevertheless, $\mu^f(\theta) \geq \mu^*(\theta)$ for all $\theta \in \Theta$.

Of course, both the naive and the optimal mechanisms are sensitive to $\mathcal{K}$. As an extreme example, if there is only one feasible system, both mechanisms will generate the same expected organizational net value. In this example, the expected organizational net value generated by a profit center is again smaller than that generated by either the optimal mechanism or the naive mechanism. For the continuous case in Example 3.1, the expected organizational net value is 5.12 even when $\xi = 0.5$. By comparison, the discreteness of $\mathcal{K}$ actually increases the expected organizational net value when the IS department is organized as a profit center. In the continuous case, the capacity of the IS department's system choice is too small compared with that which maximizes the organizational net value. In this example, the smallest feasible system has a capacity of 2. So even when θ is large, the IS department's ability to choose a smaller system in order to maximize its profit is limited. As a result, the discreteness of $\mathcal{K}$ aligns the

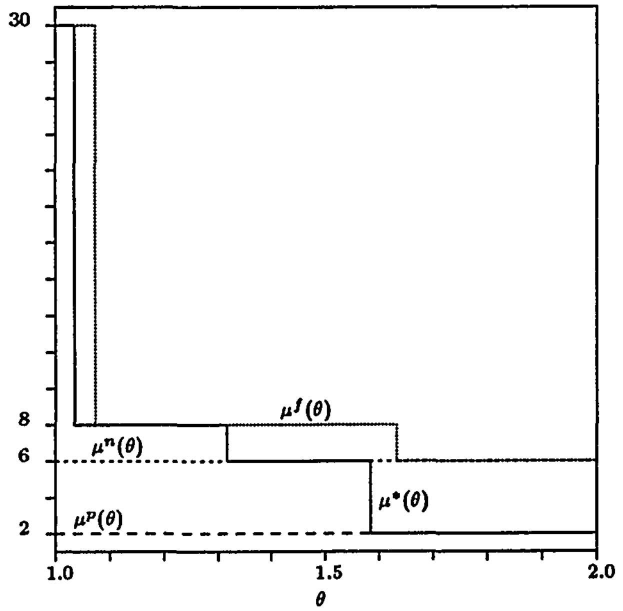


FIGURE 4.1: EXAMPLE 4.1—THE COMPARISON OF EX POST CAPACITY OF THE FOUR CASES CONSIDERED AS A FUNCTION OF θ.

IS department's decisions more closely with the decisions that maximize the expected organizational net value. Thus, even when the IS department is organized as a profit center and the feasible systems are continuous, the organization may be better off if the central management is involved in the decision on setting capacity or at least puts a lower bound on the IS department's system capacity.

Example 4.2: A Multi-server Queue: $M/M/s$

In this example, I demonstrate how an information system can be controlled when it can be characterized as a multi-server queue. I retain all the assumptions made in the previous example, except that the system now is represented by the number of servers and $\xi = 0.5$. I further normalize the processing rate of each server to be 1 without loss of generality. The capacity cost is assumed to be a linear function of the number of servers, s: $C(s, \theta) = \theta s$.

Define p_n as the equilibrium probability that there are n jobs in the system with the following standard balance equations:

$$\begin{aligned}\lambda p_n &= (n+1)\mu p_{n+1}, n = 0,1,\ldots,s-1,\\ \lambda p_n &= s\mu p_{n+1}, n = s, s+1,\ldots.\end{aligned}$$

Letting $\rho \stackrel{\text{def}}{=} \frac{\lambda}{s}$ be the *utilization factor*, then for a fixed number of servers s,

$$p_n = \begin{cases} \frac{p_0\lambda^n}{n!} & \text{for } n \le s \\ \frac{p_0\rho^n s^s}{s!} & \text{for } n > s, \end{cases}$$

where

$$p_0 = \left[\frac{\lambda^s}{(1-\rho)s!} + \sum_{j=0}^{s-1}\frac{\lambda^j}{j!}\right]^{-1}.$$

Letting

$$P(W^q > 0) \stackrel{\text{def}}{=} 1 - \sum_{j=0}^{s-1} p_j = 1 - p_0 \sum_{j=0}^{s-1}\frac{\lambda^j}{j!},$$

the probability that a job finds all servers are busy when it enters the system, then the users' aggregate delay cost is:

$$\lambda W(\lambda, s) = \lambda\left[\frac{P(W^q > 0)}{s-\lambda} + 1\right].$$

The question now is whether or not the marginal analysis is valid for the $M/M/s$ case. Although it has been shown that $W(\lambda, s)$ is non-decreasing and convex in the traffic intensity, λ (Grassmann [36] and Lee and Cohen [63]), and non-increasing and convex in s (Dyer and Proll [30]), I am not able to prove analytically that $GV(s)$ is increasing and concave in s. Nevertheless, from Table 4.6, it is clear that $GV(s)$ is indeed increasing and concave in this example, and therefore Proposition 4.2 applies.

The Naive Mechanism. For the naive case (and the full-information case as well), let θ_1^n be the smallest realization such that there exists two systems:

$$\begin{aligned}\theta_1^n &= \frac{GV(s) - GV(s-1)}{s-(s-1)}\\ &= GV(s) - GV(s-1),\end{aligned}$$

where

$$GV(s) \stackrel{\text{def}}{=} \max_{\lambda} V(\lambda) - \lambda W(\lambda, s).$$

TABLE 4.6: EXAMPLE 4.2—THE GROSS ORGANIZATIONAL NET VALUE AND THE OPTIMAL ARRIVAL RATE.

s	7	6	5	4	3	2
$GV(s)$	16.39683	15.37083	14.18716	12.79598	11.11531	8.99306
$\lambda(s)$	5.39655	4.59048	3.78983	2.99730	2.21571	1.44686

TABLE 4.7: EXAMPLE 4.2—THE PARTITIONS OF THE FOUR CASES CONSIDERED.

	θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
Naive Mechanism	1.00000	1.02560	1.18367	1.39118	1.68067	2.00000	—
Optimal Mechanism	1.00000	1.01733	1.12245	1.26079	1.45378	1.74817	2.00000
Fully Informed Case	1.00000	1.02560	1.18367	1.39118	1.68067	2.00000	
Profit Center	1.00000	1.14674	2.00000	—	—	—	—

The induced partition can then be obtained by solving:

$$\theta_i^n = GV(s-i+1) - GV(s-i)$$

repeatedly until a system is reached for which the above equality fails to hold.

It can be easily verified from Table 4.6 that the systems with less than 3 or greater than 7 servers will never be chosen under the naive mechanism, i.e., $\theta_1^n = GV(7) - GV(6)$.

The Optimal Mechanism. Similarly, for the optimal mechanism, from Proposition 4.2,

$$\theta_1^* = \frac{1}{1+\xi}\left(GV(s) - GV(s-1) + \xi\underline{\theta}\right),$$

where again $\mu_1^* = 7$ for this example.

The Profit Center. For the profit center, the partition can be obtained by solving:

$$\theta_i^p = R(i) - R(i+1)$$

TABLE 4.8: EXAMPLE 4.2—THE EXCESS BUDGET ALLOCATION THAT THE IS MANAGER CAN OBTAIN OVER EACH INTERVAL UNDER THE NAIVE MECHANISM.

Θ_i^n	$S^n(\theta_1^n;\theta)$	$S^n(\theta_2^n;\theta)$	$S^n(\theta_3^n;\theta)$	$S^n(\theta_4^n;\theta)$	$S^n(\theta_5^n;\theta)$
$[\theta_0^n,\theta_1^n]$	[0.17920, 0.00000]	[1.10202, 0.94842]	[1.95590, 1.82790]	[2.72268, 2.62028]	[3.00000, 2.92320]
$(\theta_1^n,\theta_2^n]$	—	(0.94842, 0.00000]	(1.82790, 1.03755]	(2.62028, 1.98800]	(2.92320, 2.44899]
$(\theta_2^n,\theta_3^n]$	—	—	(1.03755, 0.00000]	(1.98800, 1.15796]	(2.44899, 1.82646]
$(\theta_3^n,\theta_4^n]$	—	—	—	(1.15796, 0.00000]	(1.82646, 0.95799]
$(\theta_4^n,\theta_5^n]$	—	—	—	—	(0.95799, 0.00000]

TABLE 4.9: EXAMPLE 4.2—THE RESULTING EX POST NUMBER OF SERVERS.

Interval	Θ_1	Θ_2	Θ_3	Θ_4	Θ_5	Θ_6	$E\{NV(\theta)\}$
Naive Mechanism	3	3	3	3	3	—	5.11531
Optimal Mechanism	7	6	5	4	3	2	6.57872
Fully Informed Case	7	6	5	4	3	—	6.94068
Profit Center	2	1	—	—	—	—	4.56888

repetitively provided that $\theta_i^p \in \Theta$. The profit maximizing partition and the corresponding capacity are given in Tables 4.7 and 4.9.

Discussion. Table 4.7 gives the partitions generated for the four cases considered. It is clear that, when the IS department is organized as a profit center, the IS department will always choose the system with a single server unless $\theta < 1.14674$. The ex post numbers of servers and the expected organizational net value for the four cases considered are shown in Table 4.9. Due to the effect of the informational rent, the system with 2 servers is included in the optimal mechanism, since this system is organizationally preferable when the cost is sufficiently high. When the IS department is organized as a profit center, the potential systems that it may acquire will never be chosen for the full-information case.

From Figure 4.2, $\Theta_i^f \cap \Theta_i^* \neq \emptyset$ for all $i \in K^f$. That is, the system capacity is not distorted over $(\theta_i^f, \theta_i^*]$ for all $i \in K^f$. This again illustrates how the discreteness of the feasible systems constrains the central management's ability to control the IS department. Note that $\frac{\theta_i^* - \theta_{i-1}^f}{\theta_i^f - \theta_{i-1}^f}$ is the fraction of range Θ_i^f over which the decision is not distorted. It is clear from Figure 4.2 that $\frac{\theta_i^* - \theta_{i-1}^f}{\theta_i^f - \theta_{i-1}^f}$ is decreasing in i for all $i \in \{1, \ldots, s-1\}$. Thus, for large i, the possibility of system distortion is higher.

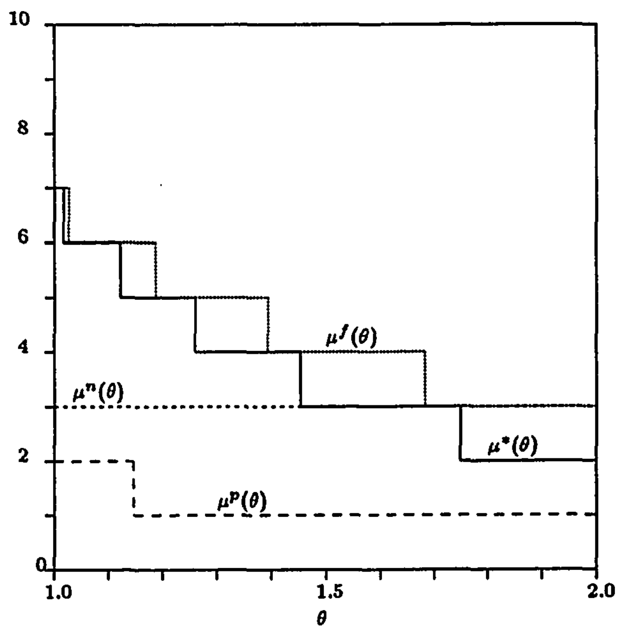


FIGURE 4.2: EXAMPLE 4.2—THE COMPARISON OF EX POST CAPACITY OF THE FOUR CASES CONSIDERED AS A FUNCTION OF θ.

Chapter 5

Limited Communication

5.1 Introduction

As discussed in Chapter 1, there are many sources that can limit communication among members of an organization. Since unlimited and costless communication is required for the revelation principle to be valid, the performance of a decentralized profit center may not be replicated by the optimal centralized mechanism if the communication between the central management and the IS department is sufficiently limited. Even when unlimited communication is possible, it may be so expensive as to be suboptimal. Thus, "even well-intentioned members of an organization ... may have trouble communicating all the information they possess to their relevant co-members, because it is too time consuming or because the information is hard to 'codify' to make it understandable to its receivers. Thus, decisions that would be profit maximizing under full communication will not be made under imperfect communication" (Tirole [106], p. 49). Then the question is: When should an organization centralize or decentralize the IS-related decisions, given that the IS operations are highly specialized? Although my model is highly stylized and only serves as a crude abstraction of reality, my results may nevertheless generate some testable implications, subject to empirical evaluation. One possible question is: Should IS-related decisions tend to be centralized as an organization's communication system improves or as the systems become more user-friendly? To simplify the exposition, the set of feasible systems is assumed to be continuous, as in Chapter 3.

5.2 Model of Limited Communication

One relevant concept related to limited communication and knowledge is that of "bounded rationality." When human agents are subject to bounded rationality, their behavior is "*intendedly* rational, but only *limitedly* so" (Simon [100]; Williamson [116]). The potential problems associated with limited communication have been recognized in areas of Coordination Theory and Distributed Artificial Intelligence (Genesereth [33]; Georgeff [34, 35]; Rosenschein [94]). Such limited communication and knowledge phenomena are thus likely to be the rule rather than the exception. Due to the deficiency of required knowledge, the ability of the members of an organization to make effective decisions will be constrained in a similar way. That is, it is more reasonable that the central management's ability to serve as an effective mechanism designer should also be limited for the problem that I study.

To model formally the limited communication that occurs in a real-world environment is intractable. There are many models of limited communication existing in the economics literature (see, e.g., Green and Laffont [38, 39, 40]; Oniki [90]). Due to the complexity of human communication, these models inevitably rely upon ad hoc specifications. The prevailing phenomenon of specialization in organizations complicates the problem even further. Because of specialization, the "private information" of the members of an organization will appear as "know-how" instead of some plain "state information." As a result, one member's knowledge may not be able to be communicated without incurring prohibitively high costs.

In this chapter I do not intend to provide a more general model of limited communication and knowledge. Instead, I adopt the model of limited communication provided in Melumad et al. [71] to model limited communication as well as limited knowledge. This approach is reasonable because the central management should not be able to prescribe effective decisions based on a message sent by the IS manager if it cannot understand the message perfectly. To formalize the central management's limitation, I assume that it can only construct a mechanism that can be supported by a "limited communication system." A limited communication system does not need to be a physical device; it may merely reflect the extent of the knowledge possessed by the central management about the IS operations. In other words, due to the complexity of the IS operations, the central management cannot prescribe a comprehensive budget allocation scheme parameterized by every possible realization of θ, as in the previous chapters. A limited communication system, then, constrains both the central management's ability to communicate and its ability to design a mechanism. Hereafter I use the term "limited communication" to represent both types of constraints and focus on formalizing a limited communication system. As in Melumad et al. [71], I model a limited communication system by a finite message set.

Kofman and Ratliff [60] take a similar approach to construct a model of bounded rationality and costly communication. A bit-by-bit communication algorithm is used to formalize a limited and costly communication device for exchange of information and coordination in a two-person team setting. They identify the sufficient conditions under which attention can be confined to convex partitions of the support of random variables. Depending on the number of bits that can be used for communication, the resulting convex partitions consist of an immediate action cell and one or more than one message cells. In their model, a communication algorithm dictates the communication process as well as the choice of actions. An algorithm may terminate and actions may be taken before the capacity of the algorithm is reached (the available bits are exhausted). Based on this model, they compare the effectiveness of monolog and dialog algorithms. The cost of communication is measured by the communication length in the worst case, or the number of bits in the longest possible information exchange. In so doing, the number of bits can be used to represent the "communication budget" of a communication algorithm.

Since the IS manager is the only party in my model with private information and the central management is the only party making IS-related decisions, my model of limited communication systems can be considered to be a monolog algorithm with a fixed communication length and decisions being made by the central management at termination of the algorithm. Thus if b is the number of bits that a communication system is designed to support, this system can distinguish 2^b different messages. For example, if the message set consists of two messages: {high, low}, "0" can represent "low" and "1" can represent "high." Since I am studying an environment with incentive conflicts, the issue of incentive compatibility needs to be considered when designing a communication system.

5.3 Centralized Mechanisms

In the previous chapters, I rely upon the revelation principle to derive the centralized mechanism, which is not dominated by any other method of control. However, the revelation principle can fail in the presence of limited communication. As a result, an optimally designed centralized mechanism may not be able to replicate the performance of a decentralized mechanism like a profit center. Yet if we restrict our attention to centralized mechanisms, following the argument used to establish the revelation principle, we can again without loss of generality focus on mechanisms for which the IS manager obeys the provided communication rule $\sigma(\cdot)$ (Melumad et al. [71]). Consequently, when dealing with centralized mechanisms, we can further restrict our attention to the direct revelation mechanisms defined by the following three outcome functions:

$$\begin{aligned} \lambda : \mathcal{M} &\rightarrow \mathcal{R}_+ \\ \mu : \mathcal{M} &\rightarrow \mathcal{R}_+ \\ T : \mathcal{M} &\rightarrow \mathcal{R}. \end{aligned}$$

A centralized mechanism then specifies the decisions $(\lambda(m), \mu(m))$ to be implemented by the IS department through the transfer $T(m)$ for all possible messages $m \in \mathcal{M}$. A mechanism is compatible with $\mathcal{M}$ if and only if for all $m \in \mathcal{M}$, $A \subseteq \Theta_m, \tilde{\theta}, \hat{\theta} \in A$ implies that $\Gamma(\tilde{\theta}) = \Gamma(\hat{\theta})$. In other words, a mechanism is compatible with $\mathcal{M}$ if and only if the central management's decisions are the same for all $\theta \in \Theta_m$ and $m \in \mathcal{M}$. A mechanism with a finite message set $\mathcal{M}$ is incentive compatible if and only if

$$T(\sigma(\theta)) - C(\mu(\sigma(\theta)), \theta) \geq T(m) - C(\mu(m), \theta), \qquad \forall \theta \in \Theta, \forall m \in \mathcal{M}.$$

We first show that there is no loss of generality in restricting attention to interval partitions induced by the message set $\mathcal{M}$.

LEMMA 5.1 *If there are two cost parameters $\theta_j, \theta_{j+1} \in \Theta$ such that $\theta_j < \theta_{j+1}$ and $\sigma(\theta_j) = \sigma(\theta_{j+1}) = m$, then the incentive compatibility constraints (5.1) imply that $\mu(\tilde{m}) = \mu(m)$ and $T(\tilde{m}) = T(m)$, where $\tilde{m} = \sigma(\theta)$ for some arbitrary $\theta \in (\theta_j, \theta_{j+1})$. Then without loss of generality the IS manager's communication rule can be recast as $\sigma(\theta) = m$ for all $\theta \in [\theta_j, \theta_{j+1}]$.*

PROOF. Suppose there is a $\tilde{\theta} \in (\theta_j, \theta_{j+1})$ such that $\sigma(\tilde{\theta}) = \tilde{m} \neq m$. Then (5.1) implies that

$$\begin{aligned} T(\tilde{m}) - C(\mu(\tilde{m}), \tilde{\theta}) &\geq T(m) - C(\mu(m), \tilde{\theta}) \\ T(m) - C(\mu(m), \theta_j) &\geq T(\tilde{m}) - C(\mu(\tilde{m}), \theta_j) \\ T(m) - C(\mu(m), \theta_{j+1}) &\geq T(\tilde{m}) - C(\mu(\tilde{m}), \theta_{j+1}), \end{aligned}$$

or

$$\begin{aligned} T(m) - T(\tilde{m}) &\leq C(\mu(m), \tilde{\theta}) - C(\mu(\tilde{m}), \tilde{\theta}) \\ T(m) - T(\tilde{m}) &\geq C(\mu(m), \theta_j) - C(\mu(\tilde{m}), \theta_j) \\ T(m) - T(\tilde{m}) &\geq C(\mu(m), \theta_{j+1}) - C(\mu(\tilde{m}), \theta_{j+1}). \end{aligned}$$

The first two inequalities imply that

$$C(\mu(\tilde{m}), \theta_j) - C(\mu(m), \theta_j) \geq C(\mu(\tilde{m}), \tilde{\theta}) - C(\mu(m), \tilde{\theta}).$$

By assumptions that $C_{\mu\theta} > 0$ and $\tilde{\theta} > \theta_j$, $\mu(\tilde{m}) \leq \mu(m)$. Similarly, the first and the third inequalities imply that

$$C(\mu(\tilde{m}), \theta_{j+1}) - C(\mu(m), \theta_{j+1}) \geq C(\mu(\tilde{m}), \tilde{\theta}) - C(\mu(m), \tilde{\theta})$$

which in turn implies that $\mu(\tilde{m}) \geq \mu(m)$, since $\tilde{\theta} < \theta_{j+1}$. Thus $\mu(\tilde{m}) = \mu(m)$ and therefore $T(\tilde{m}) = T(m)$. Since the IS manager will be indifferent between messages $\tilde{m}$ and m, the message $\tilde{m}$ is redundant and thus can be deleted without loss of generality.||

From Lemma 5.1, the set of cost parameters corresponding to a particular message must be a convex set and thus an interval. Consequently, the message space can be represented as a partition over Θ. This characterization of the message space is natural. For instance, as a simple example, the message set may consist of only two elements: {high, low}, and the message "low" ("high") can be viewed as corresponding to the condition that the IS department's cost parameter be within $[\underline{\theta}, \theta_1]$ ($(\theta_1, \overline{\theta}]$).

Since a particular interval corresponds to a single message, the decisions must the same for the whole interval. Thus in my model a limited communication system can be characterized by a finite set of intervals that partitions Θ. After deleting redundant messages, I can focus on the minimal message set that implements the outcomes implementable with the original message set. Each interval within the partition induced by this minimal message set can be viewed as a possible "message." Then a mechanism $(\lambda(\cdot), \mu(\cdot), T(\cdot))$ is incentive compatible if and only if

$$T(m) - C(\mu(m), \theta) \geq T(n) - C(\mu(n), \theta), \qquad \forall \theta \in \Theta_m, \forall m, n \in \mathcal{M}. \tag{5.1}$$

To define a feasible mechanism, an additional set of constraints is again needed that enables the IS department at least to balance its budget if the IS manager obeys the communication rule, i.e.,

$$T(m) - C(\mu(m), \theta) \geq 0, \qquad \forall \theta \in \Theta_m, \forall m \in \mathcal{M}. \tag{5.2}$$

Formally, let $\mathcal{M}$ denote a finite message set with $|\mathcal{M}| = M$, the cardinality of $\mathcal{M}$, and let $COM(M)$ denote a communication system supported with M messages. That is, a $COM(M)$ is capable of supporting any interval partition up to M intervals. Since $COM(1)$ contains only one message, it corresponds to the "null" communication case, where the central management has an extremely coarse knowledge about IS operations. As $M \to \infty$, $COM(M)$ corresponds to the ordinary asymmetric information case with unlimited communication. The IS manager is provided with a communication rule $\sigma : \Theta \to \mathcal{M}$.

Let the corresponding intervals of the partition induced by a $COM(M)$ be denoted by $\Theta_1 \equiv [\theta_0, \theta_1]$ and $\Theta_m \equiv (\theta_{m-1}, \theta_m]$ for all $m \in \{2, \ldots, M\}$, where $\theta_0 = \underline{\theta}$ and $\theta_M = \overline{\theta}$. Θ_m denotes message m. Let $\Theta_m(\theta)$ be the interval within which the true parameter

θ lies. With a limited communication system, even though the IS manager intends to report truthfully (so that $\hat{\theta}(\theta) = \theta$), the central management can only receive the message m (where $\theta \in \Theta_m(\theta)$). Then incentive compatibility here simply requires the IS manager not to pick a message n such that $\theta \notin \Theta_n$. Again, for the case with $\mathcal{M} = \{\text{high}, \text{low}\}$, incentive compatibility requires the IS manager not to report the department's cost as "high" if $\theta \in \Theta_1$, and vice versa. If bit-by-bit communication is used to support interval partitions, then $\lceil \log_2 M \rceil$ bits are needed to distinguish M possible messages.

Given a fixed $COM(M)$ system, the program for deriving the maximum expected organizational net value can be recast as:

$$\max_{\lambda(\cdot),\mu(\cdot),T(\cdot)} \sum_{m=1}^{M} \int_{\Theta_m} NV(\lambda(m),\mu(m),T(m),\theta)\, dF(\theta) \tag{5.3}$$

subject to

$$T(m) - C(\mu(m),\theta) \geq T(n) - C(\mu(n),\theta), \qquad \forall\theta \in \Theta_m, \forall m,n \in \{1,\ldots,M\} \tag{5.4}$$

$$T(m) - C(\mu(m),\theta) \geq 0, \qquad \forall\theta \in \Theta_m, \forall m \in \{1,\ldots,M\}, \tag{5.5}$$

where the ex post organizational net value

$$NV(\lambda(m),\mu(m),T(m),\theta) \stackrel{\text{def}}{=}$$
$$V(\lambda(m)) - \lambda(m)W(\lambda(m),\mu(m)) - C(\mu(m),\theta) - \xi\,[T(m) - C(\mu(m),\theta)].$$

The following lemma characterizes the set of feasible mechanisms.

LEMMA 5.2 *Given an arbitrary partition: $\{\Theta_m : m = 1,\ldots,M\}$, if $\mu(m)$ is decreasing, then the IS manager's communication rule is equivalent to requiring $\sigma(\theta) = \theta_m$ for all $\theta \in \Theta_m$ and all $m \in \mathcal{M}$, and the budget allocation is incentive compatible if and only if for all $m \in \{1,2,\ldots,M\}$,*

$$T(m) = C(\mu(m),\theta_m) + \sum_{j=m+1}^{M} [C(\mu(j),\theta_j) - C(\mu(j),\theta_{j-1})], \tag{5.6}$$

where by convention

$$\sum_{j=m+1}^{M} [C(\mu(j),\theta_j) - C(\mu(j),\theta_{j-1})] \equiv 0.$$

Moreover, if

$$T(M) = C(\mu(M), \theta_M), \tag{5.7}$$

then (5.5) is satisfied for all $\theta \in \Theta$.

PROOF. Given an arbitrary message space: $\{\Theta_m; m = 1, \dots, M\}$, and assuming that the central management's decision on the system capacity $\mu(m)$ is decreasing, then for any $m < M$, local incentive compatibility requires, for all $\theta \in \Theta_m$,

$$T(m) - C(\mu(m), \theta) \geq T(m+1) - C(\mu(m+1), \theta), \tag{5.8}$$

and for all $\theta \in \Theta_{m+1}$

$$T(m+1) - C(\mu(m+1), \theta) \geq T(m) - C(\mu(m), \theta). \tag{5.9}$$

Since by assumption $\mu(m)$ is decreasing and $C_{\mu\theta} > 0$, (5.8) and (5.9) imply that

$$T(m) = T(m+1) + C(\mu(m), \theta_m) - C(\mu(m+1), \theta_m), \tag{5.10}$$

for all $\theta \in \Theta_m$. Hence the IS manager's communication rule is equivalent to $\sigma(\theta) = \theta_m$ for all $\theta \in \Theta_m$ and all $m \in \mathcal{M}$.

Therefore, for any $m < M$, local incentive compatibility requires, for all $\theta \in \Theta_m$,

$$T(m) - C(\mu(m), \theta) \geq T(m+1) - C(\mu(m+1), \theta), \tag{5.11}$$

and for all $\theta \in \Theta_{m+1}$,

$$T(m+1) - C(\mu(m+1), \theta) \geq T(m) - C(\mu(m), \theta), \tag{5.12}$$

and so

$$T(m) = T(m+1) + C(\mu(m), \theta_m) - C(\mu(m+1), \theta_m). \tag{5.13}$$

(5.13) must also be globally incentive compatible. Assuming $\theta \in \Theta_m$,

$$T(m) - C(\mu(m), \theta) \geq T(j) - C(\mu(j), \theta), \qquad \forall j \in \{m+1, \dots, M\}.$$

If $j = m + 1$, we are done by local incentive compatibility. Let $j = m + 2$. Since from the IS manager's standpoint, claiming $\theta \in \Theta_m$ is at least as good as claiming $\theta \in \Theta_{m+1}$, it is sufficient to show

$$T(m+1) - C(\mu(m+1), \theta) \geq T(m+2) - C(\mu(m+2), \theta). \tag{5.14}$$

From (5.13),

$$T(m+1) = T(m+2) + C(\mu(m+1), \theta_{m+1}) - C(\mu(m+2), \theta_{m+1}),$$

(5.14) is then equivalent to requiring

$$C(\mu(m+1), \theta_{m+1}) - C(\mu(m+1), \theta) \geq C(\mu(m+2), \theta_{m+1}) - C(\mu(m+2), \theta).$$

But by assumption $\mu(m)$ is decreasing in m and $C_{\mu\theta} > 0$, so the above inequality is satisfied. Following this process, it can be shown that (5.13) satisfies all the upward constraints. Similarly for the downward constraints, it must be shown

$$T(m) - C(\mu(m), \theta) \geq T(j) - C(\mu(j), \theta), \qquad \forall j \in \{1, \ldots, m-1\}.$$

Again if $j = m - 1$, we are done by local incentive compatibility. Let $j = m - 2$, and then from (5.13),

$$T(m-1) - C(\mu(m-1), \theta) \geq T(m-2) - C(\mu(m-2), \theta) \tag{5.15}$$

is equivalent to requiring

$$C(\mu(m-2), \theta) - C(\mu(m-1), \theta) \geq C(\mu(m-2), \theta_{m-2}) - C(\mu(m-1), \theta_{m-2}).$$

Again by assumption $\mu(m)$ is decreasing in m and $C_{\mu\theta} > 0$, so the above inequality is satisfied. Following this process, it can be shown that (5.13) also satisfies all the downward constraints. Thus (5.13) is globally incentive compatible.

From (5.13), it is easy to show by induction that for all $m \in \{1, 2, \ldots, M-1\}$,

$$T(m) = C(\mu(m), \theta_m) + \sum_{j=m+1}^{M} \left[C(\mu(j), \theta_j) - C(\mu(j), \theta_{j-1})\right].$$

Since the excess budget allocation is

$$\sum_{j=m+1}^{M} \left[C(\mu(j), \theta_j) - C(\mu(j), \theta_{j-1})\right],$$

setting $T(M) = C(\mu(M), \theta_M)$ will satisfy the budget constraint for all $\theta \in \Theta$. ||

Given that the budget allocation rule satisfies (5.4) and the capacity decision $\mu(m)$ is non-increasing, the IS manager will send a message according to the communication rule that she is supposed to send. To derive the optimal mechanism for a fixed $COM(M)$ system, one only need to seek an undominated centralized mechanism in which $\mu(m)$ is non-increasing.

Since the budget allocation $T(m)$ only depends on $\mu(\cdot)$, it can be eliminated from the central management's objective function, and consequently the mechanism only involves two outcome functions $(\lambda(\cdot), \mu(\cdot))$. Fixing $(\lambda(m), \mu(m))$ and substituting (5.6) for $T(m)$ in the central management's objective function yields:

$$\begin{aligned} E\{NV(\theta); M\} &\stackrel{\text{def}}{=} \sum_{m=1}^{M} \int_{\Theta_m} \Big\{V(\lambda(m)) - \lambda(m)W(\lambda(m), \mu(m)) - (1-\xi)C(\mu(m), \theta) \\ &\qquad -C(\mu(m), \theta_m) - \sum_{j=m+1}^{M} [C(\mu(j), \theta_j) - C(\mu(j), \theta_{j-1})]\Big\} dF(\theta) \\ &= \sum_{m=1}^{M} \Big\{V(\lambda(m)) - \lambda(m)W(\lambda(m), \mu(m)) - \frac{(1-\xi)\int_{\theta_{m-1}}^{\theta_m} C(\mu(m), \theta)\, dF(\theta)}{\int_{\theta_{m-1}}^{\theta_m} dF(\theta)} \\ &\qquad -\xi C(\mu(m), \theta_m) - \xi \sum_{j=m+1}^{M} [C(\mu(j), \theta_j) - C(\mu(j), \theta_{j-1})]\Big\} \int_{\theta_{m-1}}^{\theta_m} dF(\theta). \end{aligned}$$

By convention,

$$\sum_{j=m+1}^{M} [C(\mu(j), \theta_j) - C(\mu(j), \theta_{j-1})] \equiv 0,$$

and then by induction:

$$\begin{aligned} &\sum_{m=1}^{M} \sum_{j=m+1}^{M} [C(\mu(j), \theta_j) - C(\mu(j), \theta_{j-1})]\, [F(\theta_m) - F(\theta_{m-1})] = \\ &\qquad \sum_{m=1}^{M} [C(\mu(m), \theta_m) - C(\mu(m), \theta_{m-1})]\, F(\theta_{m-1}). \end{aligned}$$

Since

$$\begin{aligned} &C(\mu(m), \theta_m) + [C(\mu(m), \theta_m) - C(\mu(m), \theta_{m-1})] \frac{F(\theta_{m-1})}{F(\theta_m) - F(\theta_{m-1})} = \\ &\qquad \frac{\int_{\theta_{m-1}}^{\theta_m} \{C(\mu(m), \theta) + \beta(\theta) C_\theta(\mu(m), \theta)\} dF(\theta)}{\int_{\theta_{m-1}}^{\theta_m} dF(\theta)}, \end{aligned}$$

(5.3) can be recast as:

$$\max_{\lambda(\cdot), \mu(\cdot)} \sum_{m=1}^{M} H(\lambda(m), \mu(m); \theta_m, \theta_{m-1})\, [F(\theta_m) - F(\theta_{m-1})]\,, \tag{5.16}$$

where

$$H(\lambda(m),\mu(m);\theta_m,\theta_{m-1}) \stackrel{\text{def}}{=} V(\lambda(m)) - \lambda(m)W(\lambda(m),\mu(m)) - VC(\mu(m),\theta_m,\theta_{m-1})$$

is the virtual organizational net value,

$$VC(\mu(m),\theta_m,\theta_{m-1}) \stackrel{\text{def}}{=} \frac{\int_{\theta_{m-1}}^{\theta_m} \{C(\mu(m),\theta) + \xi\beta(\theta)C_\theta(\mu(m),\theta)\}\, dF(\theta)}{\int_{\theta_{m-1}}^{\theta_m} dF(\theta)}$$

is the virtual capacity cost, and $\beta(\theta) \equiv \frac{F(\theta)}{f(\theta)}$.

In the presence of information asymmetry, regardless of whether the communication is limited or not, the central management will distort the capacity (and thereby the price and the budget allocation) so as to make a tradeoff between operational efficiency and the IS manager's informational rent. But with limited communication, the resulting incentive compatible mechanism is discrete, so the effect of limited communication is similar to the effect of the finite set of feasible systems studied in Chapter 4. However, here the effective capacity can still be adjusted across intervals continuously before the optimal centralized mechanism is reached (although the capacity must be the same within the same interval). Letting $\mu^*(\theta)$ and $\mu^*(m)$ be the optimal capacity for the continuous case and the limited communication case, respectively, then for every m there must exist some $\theta \in \Theta_m$ such that $\mu^*(\theta) = \mu^*(m)$, provided that $C(\mu,\theta)+\xi\beta(\theta)C_\theta(\mu,\theta)$ is monotone in θ. This phenomenon can be seen in Figures 5.3 and 5.4.

Given θ, Lemma 5.2 implies that the IS manager's expected ex post informational rent is:

$$\xi \int_{\Theta_m} \left\{ [C(\mu(m),\theta_m) - C(\mu(m),\theta)] + \sum_{j=m+1}^{M} [C(\mu(j),\theta_j) - C(\mu(j),\theta_{j-1})] \right\} dF(\theta). \tag{5.17}$$

However, the IS manager's virtual informational rent in (5.16) is evaluated as:

$$\xi \int_{\Theta_m} \left\{ C(\mu(m),\theta_m) - C(\mu(m),\theta) + \frac{[C(\mu(m),\theta_m) - C(\mu(m),\theta_{m-1})]F(\theta_{m-1})}{F(\theta_m) - F(\theta_{m-1})} \right\} dF(\theta). \tag{5.18}$$

Comparing (5.17) with (5.18), when $m = 1$, the IS manager's ex post informational rent is greatest, but the mechanism only takes into account

$$\xi \int_{\Theta_1} \{C(\mu(1),\theta_1) - C(\mu(1),\theta)\} dF(\theta)$$

(since $F(\theta_0) = 0$). As a result, the extent of distortion to the decisions is also smallest even though the IS manager can earn the largest informational rent ex post. On the contrary, when $m = M$, the IS manager's ex post informational rent is smallest, but the organization's objective function takes full account of the IS manager's virtual informational rent. Note that, due to the effect of limited communication, the capacity will still be distorted unless $\theta = \theta_1$ because the cost parameter is evaluated at θ_1 instead of $\underline{\theta}$. Thus the usual "no distortion at the bottom" property holds if θ_1 is considered the "bottom" of the message set. Similarly, the IS manager's ex post informational rent is strictly positive unless $\theta = \theta_M = \overline{\theta}$.

In Lemma 5.2, I rely upon the assumption that $\mu(m)$ is decreasing. This condition in general is not satisfied for arbitrary probability distribution functions and capacity cost functions. In order for (5.6) to be incentive compatible, some restrictions on the functions involved are needed. The sufficient condition for (5.6) to be incentive compatible is given in the following proposition.

PROPOSITION 5.1 *Suppose that for all m the Hessian matrix of $H(\lambda, \mu, \theta_m, \theta_{m-1})$ with respect to λ and μ is negative definite and that, for a fixed μ,*

$$C_\mu(\mu, \theta) + \xi\beta(\theta)C_{\theta\mu}(\mu, \theta) \tag{5.19}$$

is increasing in θ, there exists for any partition, $\{\Theta_m : m = 1, \ldots, M\}$, a unique optimal incentive compatible mechanism:

$$\{(\lambda^c(\theta_m, \theta_{m-1}), \mu^c(\theta_m, \theta_{m-1})) : m = 1, \ldots, M\},$$

which is the solution to the following equations: For $m = 1, \ldots, M$,

$$0 = V'(\lambda) - W(\lambda, \mu) - \lambda W_\lambda(\lambda, \mu); \tag{5.20}$$

$$0 = -\lambda W_\mu(\lambda, \mu) - VC_\mu(\mu, \theta_m, \theta_{m-1}). \tag{5.21}$$

PROOF. The equations (5.20) and (5.21) are the first-order conditions of (5.16) with respect to λ and μ for each m. Given the Hessian matrix of $H(\lambda, \mu, \theta_m, \theta_{m-1})$ with respect to λ and μ is negative definite, it is easy to show that by the implicit function theorem

$$\begin{aligned} \operatorname{sgn}\left(\mu^c_{\theta_m}\right) &= -\operatorname{sgn}\left(VC_{\mu\theta_m}\right) \\ \operatorname{sgn}\left(\mu^c_{\theta_{m-1}}\right) &= -\operatorname{sgn}\left(VC_{\mu\theta_{m-1}}\right). \end{aligned}$$

But by assumption that

$$C_\mu(\mu,\theta) + \xi\beta(\theta)C_{\theta\mu}(\mu,\theta)$$

is increasing in θ, $VC_\mu(\mu,\theta_m,\theta_{m-1})$ is increasing in both θ_m and θ_{m-1}. So $\mu^c(\theta_m,\theta_{m-1})$ is decreasing in m. ||

Clearly, as given in Melumad et al. [71], if $C(\mu,\theta) = \tau(\theta)c(\mu)$ with $\tau(\cdot)$ and $c(\cdot)$ increasing, $\tau(\cdot)$ convex, and $\beta(\theta)$ increasing, my assumption reduces to requiring that

$$\tau(\theta)c'(\mu) + \beta(\theta)\tau'(\theta)c'(\mu)$$

is increasing in θ, which is satisfied if both $\tau'(\theta)$ and $\beta(\theta)$ are increasing. When the capacity cost function is multiplicatively separable, the organization's *virtual* objective function for each interval becomes much simpler:

$$H(\lambda,\mu;\theta_m,\theta_{m-1}) = V(\lambda) - \lambda W(\lambda,\mu) - \gamma(\theta_m,\theta_{m-1})c(\mu),$$

where

$$\gamma(\theta_m,\theta_{m-1}) = \frac{\int_{\Theta_m}[\tau(\theta) + \xi\beta(\theta)\tau'(\theta)]\,dF(\theta)}{\int_{\Theta_m} dF(\theta)}.$$

Thus, when the capacity cost function is multiplicatively separable, if $\tau(\theta) + \tau'(\theta)\beta(\theta)$ is increasing, then the optimal capacity is decreasing. Hence, when $\tau(\theta) = \theta$, in order for $\mu^c(\theta_m,\theta_{m-1})$ to be decreasing in m, it is sufficient to have $\beta'(\theta) > -1$, which can be shown to be exactly required for the unlimited communication case to be globally incentive compatible.

Note that the structure of the partition will affect the IS department's informational rent, and therefore the way that Θ is partitioned can affect the expected organizational net value. Furthermore, the expected organizational net value is not necessarily monotone in M unless a system with a larger message set generates a finer partition. That is, a $COM(M)$ system can generate an expected organizational net value that is at least as high as that which can be generated by a $COM(X)$ system, where $X < M$, if for all $m = 1, \ldots, M$, there exists a $\Theta_t \in \{\Theta_j : j = 1, \ldots, X\}$ such that $\Theta_m \subseteq \Theta_t$, since the central management can always ignore the extra messages provided by a system that can induce a finer partition.

Limited communication will clearly have effects on the expected organizational net value only if the IS department is governed by a centralized mechanism. When the IS department is organized as a profit center, since the decisions are delegated to the IS manager, the expected organizational net value will not be affected. Although by evaluating the IS department as a profit center, the central management loses the control of the IS department's decisions, the "flexibility gain" from delegating decisions to the better-informed IS manager may outweigh the "control loss" with limited communication. As an extreme example, if $M = 1$, $E\{NV^c; 1\}$ is independent of $\underline{\theta}$, but $E\{NV^p\}$ is a decreasing function of $\underline{\theta}$. As a result, if $\underline{\theta}$ is sufficiently low, a profit center may outperform a centralized mechanism. Thus, with limited communication, the revelation principle can fail even without accounting for the communication costs.

5.4 Example 5.1

I continue to make use of Example 3.1 to examine the effect of a *COM*(2) system.

The Profit Center. Given the above assumptions, it is easy to show that the expected organizational net value is:

$$\begin{aligned} E\{NV^p(\theta)\} &\equiv \int_\Theta \{V(\lambda^p(\theta)) - \lambda^p(\theta)W(\lambda^p(\theta), \mu^p(\theta)) - C(\mu^p(\theta), \theta) - \xi\pi(\theta)\}\, dF(\theta) \\ &= \left[\frac{(3-\xi)k^2 \ln\theta}{4(\bar{\theta}-\underline{\theta})} - \frac{2(2-\xi)k\sqrt{\theta} - (1-\xi)\theta}{(\bar{\theta}-\underline{\theta})}\right]_{\underline{\theta}}^{\bar{\theta}}, \end{aligned}$$

where

$$\begin{aligned} \lambda^p(\theta) &= \left[\frac{k - 2\sqrt{\theta}}{2\theta}\right]^2 \\ \mu^p(\theta) &= \frac{k\left[k - 2\sqrt{\theta}\right]}{4\theta_2} \end{aligned}$$

and $\pi(\theta)$ are the profit center's optimal decisions and the corresponding profit, respectively. When $\xi = 1$, all the cost terms cancel out so that

$$NV^p(\theta) = V(\lambda^p(\theta)) - \lambda^p(\theta)V'(\lambda^p(\theta)),$$

which is positive provided that $V(\lambda)$ is concave.

The Centralized Mechanism. From Lemma 5.2, the incentive compatible budget allocations are:

$$\begin{aligned} T(1) &= \theta_1\,\mu(\theta_1,\underline{\theta}) + (\overline{\theta} - \theta_1)\,\mu(\overline{\theta},\theta_1) \\ T(2) &= \overline{\theta}\,\mu(\overline{\theta},\theta_1), \end{aligned}$$

and then the excess budget allocation equals:

$$(\theta_m - \theta)\,\mu(\theta_m,\theta_{m-1}) + (\overline{\theta} - \theta_1)\,\mu(\overline{\theta},\theta_1)$$

if $\theta \in \Theta_1$, and equals

$$(\overline{\theta} - \theta)\,\mu(\overline{\theta},\theta_1),$$

otherwise. Then, for $m = 1$ and 2,

$$\begin{aligned} &H(\lambda^c(\theta_m,\theta_{m-1}),\mu^c(\theta_m,\theta_{m-1});\theta_m,\theta_{m-1}) = V(\lambda^c(\theta_m,\theta_{m-1})) \\ &\quad -\lambda^c(\theta_m,\theta_{m-1})W(\lambda^c(\theta_m,\theta_{m-1}),\mu^c(\theta_m,\theta_{m-1})) - \gamma^c(\theta_m,\theta_{m-1})\mu^c(\theta_m,\theta_{m-1}), \end{aligned}$$

where

$$\begin{aligned} \lambda^c(\theta_m,\theta_{m-1}) &= \left[\frac{k - \sqrt{\gamma^c(\theta_m,\theta_{m-1})}}{\gamma^c(\theta_m,\theta_{m-1})}\right]^2 \\ \mu^c(\theta_m,\theta_{m-1}) &= \frac{k\left[k - \sqrt{\gamma^c(\theta_m,\theta_{m-1})}\right]}{\gamma^c(\theta_m,\theta_{m-1})^2} \\ \gamma^c(\theta_m,\theta_{m-1}) &= \frac{1+\xi}{2}(\theta_m + \theta_{m-1}) - \xi\,\underline{\theta}. \end{aligned}$$

Clearly, $\gamma^c(\theta_1,\underline{\theta}) < \gamma^c(\overline{\theta},\theta_1)$, and therefore

$$\begin{aligned} \lambda^c(\theta_1,\underline{\theta}) &> \lambda^c(\overline{\theta},\theta_1) \\ \mu^c(\theta_1,\underline{\theta}) &> \mu^c(\overline{\theta},\theta_1). \end{aligned}$$

Letting $H(\theta_m, \theta_{m-1}) \equiv H(\lambda^c, \mu^c; \theta_m, \theta_{m-1})$ yields

$$H(\theta_m, \theta_{m-1}) = \gamma^c(\theta_m, \theta_{m-1})\,\lambda^c(\theta_m, \theta_{m-1}).$$

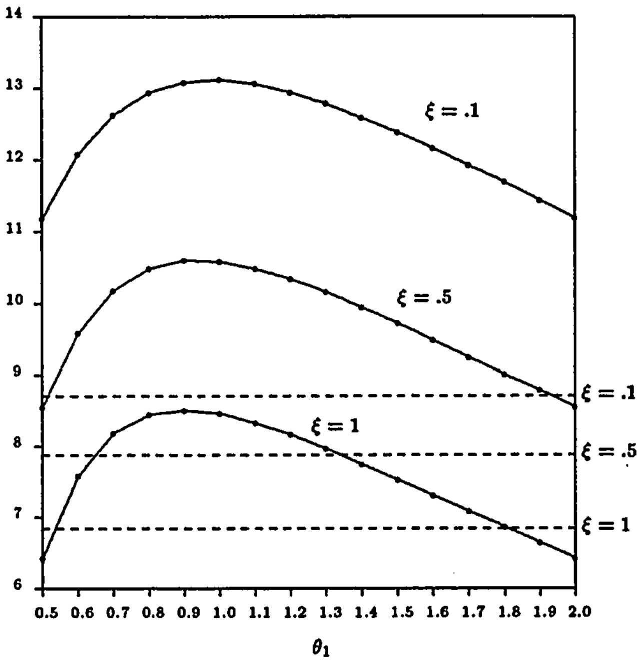
So, depending on θ_1, the expected organizational net value under the centralized mechanism is:

$$E\{NV^c; 2\} = \sum_{m=1}^{2} \gamma^c(\theta_m, \theta_{m-1})\lambda^c(\theta_m, \theta_{m-1})\frac{\theta_m - \theta_{m-1}}{\overline{\theta} - \underline{\theta}}.$$

Discussion. To give a specific example, let $k = 5$ and $\Theta = [0.5, 2]$. Figure 5.1 depicts $E\{NV^p\}$ and $E\{NV^c; 2\}$ as functions of θ_1. Since $E\{NV^p\}$ is independent of θ_1, it is a constant for every ξ. Clearly, for $\xi = 0.1$ and 0.5, the centralized mechanism dominates the profit center regardless of how Θ is partitioned. However, when $\xi = 1$, depending on θ_1, $E\{NV^c; 2\}$ can be greater than, equal to, or less than $E\{NV^p\}$. Specifically, when θ_1 is (approximately) less than 0.55 or greater than 1.8, the profit center can yield a higher expected organizational net value than the optimal centralized mechanism can. This example seems to suggest that the central management should have a stronger incentive to delegate the IS-related decisions when the IS manager has a stronger incentive problem, a counter-intuitive result. Nevertheless, with $\xi = 1$ and $\theta_1 \in (0.55, 1.8)$, the centralized mechanism can still outperform the profit center. In particular, when θ_1 is around the optimum, the centralized mechanism can outperform the profit center significantly. When θ_1 is at its optimum (which is approximately equal to 0.9), $E\{NV^c; 2\}(\approx 9)$ is greater than $E\{NV^p\}$ by about 2.

Moreover, if θ_1 is very close to either endpoint of Θ, $E\{NV^c; 2\}$ will be roughly equal to $E\{NV^c; 1\}$. For this case, a finer partition induced by a larger message set is not valuable to the organization, and a centralized mechanism can perform worse than a profit center. When $M = 1$, in order for the IS department to at least balance its budget, the central management inevitably must allocate the budget and determine capacity as if θ is always equal to $\overline{\theta}$. This case actually is identical to the case where the central management uses a maxmin strategy, and $E\{NV^c; 1\}$ will not be affected by any variation of $\underline{\theta}$ and equals 6.43. But for the profit center, since

$$\frac{d}{d\underline{\theta}}E\{NV^p\} = -NV^p(\underline{\theta}) < 0,$$



The expected organizational net value with a $COM(2)$ system when the IS department is organized as a cost and a profit center, respectively, as the cutoff point θ_1 varies from 0.5 to 2, where $k = 5$ and $\Theta = [0.5, 2]$, for cases with $\xi = 0.1, 0.5$, and 1. The solid curves are the expected organizational net values $E\{NV^c; 2\}$ under the centralized mechanism whereas the dash lines are the expected organizational net values $E\{NV^p; 2\}$ when the IS department is organized as a profit center.

FIGURE 5.1: EXAMPLE 5.1—THE EXPECTED ORGANIZATIONAL NET VALUE WITH A $COM(2)$ SYSTEM.

$E\{NV^p\}$ is strictly decreasing in $\underline{\theta}$. Consequently, the centralized mechanism should dominate the profit center when $\underline{\theta}$ is sufficiently large. For this example, when $\underline{\theta}$ is greater than approximately 0.55, the centralized mechanism can always outperform the profit center.

When decisions are delegated to the IS department and it is evaluated as a profit center, the IS manager can tailor the decisions according to the IS department's operating environment. Although the IS manager's decisions are the most efficient ones from her perspective (since they maximize her department's profit), they are not efficient at the organizational level. When $\underline{\theta}$ is sufficiently small, the "flexibility gain" outweighs the "control loss" due to the IS department's monopolistic pricing policy. But when $\underline{\theta}$ is large, the "flexibility gain" does not justify the "control loss," and consequently the profit center will fail to generate a higher expected organizational net value.

Since the way that the partition is formed has an impact on the expected organizational net value, it is then interesting to see how the optimal partition should be constructed. When the communication system is beyond the central management's control, investigating the optimal partition supportable by a finite message set is useful, since the expected net value with the optimal partition provides an upper bound for the organizational net value that can be generated by a $COM(M)$ system. For this purpose, the optimal partition compatible with a $COM(M)$ system must be derived.

5.5 Optimal Partition

Given an arbitrary partition, let $(\lambda^c(\cdot), \mu^c(\cdot))$ be the optimal mechanism with respect to that partition, and

$$H(\theta_m, \theta_{m-1}) = V(\lambda^c) - \lambda^c W(\lambda^c, \mu^c) - VC(\mu^c, \theta_m, \theta_{m-1}).$$

The maximum expected organizational net value attainable by a $COM(M)$ system then is the solution to the following program:

$$E\{NV^c; M\} = \max_{\theta_m} \sum_{m=1}^{M} H(\theta_m, \theta_{m-1}) \left[F(\theta_m) - F(\theta_{m-1})\right]. \tag{5.22}$$

subject to

$$\theta_m \geq \theta_{m-1}, \qquad \forall m \in \{1,\ldots,M-1\}. \tag{5.23}$$

For notational convenience, define

$$UV^m \stackrel{\text{def}}{=} V(\lambda^c(\theta_m,\theta_{m-1})) - \lambda^c(\theta_m,\theta_{m-1})W(\lambda^c(\theta_m,\theta_{m-1}),\mu^c(\theta_m,\theta_{m-1})).$$

The following proposition shows that, for any finite M, the optimal partition is non-degenerate and unique, provided that the assumptions given in Proposition 5.1 hold.

PROPOSITION 5.2 *Suppose that the Hessian matrix of $H(\lambda,\mu,\theta_m,\theta_{m-1})$ with respect to λ and μ is negative definite for all m and that, for a fixed μ,*

$$C_\mu(\mu,\theta) + \xi\beta(\theta)C_{\theta\mu}(\mu,\theta) \tag{5.24}$$

is increasing in θ, there exists a unique, non-degenerate optimal partition for any COM(M) with finite M, and the optimal partition is given by the following equations: For all $m \in \{1,\ldots,M-1\}$,

$$\begin{aligned} &UV^m - C(\mu^c(\theta_m,\theta_{m-1}),\theta_m) - \xi\beta(\theta_m)C_\theta(\mu^c(\theta_m,\theta_{m-1}),\theta_m) = \\ &\quad UV^{m+1} - C(\mu^c(\theta_{m+1},\theta_m),\theta_m) - \xi\beta(\theta_m)C_\theta(\mu^c(\theta_{m+1},\theta_m),\theta_m) \end{aligned} \tag{5.25}$$

with boundary conditions: $\theta_0 = \underline{\theta}$ and $\theta_M = \overline{\theta}$.

PROOF. First I show that the constraints (5.23) are non-binding; i.e, for an arbitrary interval $[\theta_{m-1},\theta_{m+1}] \subseteq \Theta$ with positive measure, there is a $\theta^*_m \in (\theta_{m-1},\theta_{m+1})$ such that

$$\theta^*_m = \arg\max_{\theta_m} \sum_{j=m}^{m+1} H(\theta_m,\theta_{m-1})[F(\theta_m) - F(\theta_{m-1})]. \tag{5.26}$$

From Proposition 5.1, $H(\theta_m,\theta_{m-1})$ is well-defined and continuously differentiable in both arguments and $\mu^c(\theta_m,\theta_{m-1})$ is decreasing in both arguments. Then, by the envelope theorem,

$$\begin{aligned} \frac{d}{d\theta_m}E\{NV^c;M\} &= f(\theta_m)[H(\theta_m,\theta_{m-1}) - H(\theta_{m+1},\theta_m)] \\ &\quad + \frac{\partial H(\theta_m,\theta_{m-1})}{\partial\theta_m}[F(\theta_m) - F(\theta_{m-1})] \end{aligned}$$

$$
\begin{aligned}
&+\frac{\partial H(\theta_{m+1},\theta_m)}{\partial\theta_m}[F(\theta_{m+1})-F(\theta_m)] \\
=\;& f(\theta_m)\Big\{UV^m - C(\mu^c(\theta_m,\theta_{m-1}),\theta_m) \\
&-\xi\beta(\theta_m)C_\theta(\mu^c(\theta_m,\theta_{m-1}),\theta_m) \\
&-[UV^{m+1} - C(\mu^c(\theta_{m+1},\theta_m),\theta_m) \\
&-\xi\beta(\theta_m)C_\theta(\mu^c(\theta_{m+1},\theta_m),\theta_m)]\Big\}. \qquad (5.27)
\end{aligned}
$$

Since

$$
\lim_{\theta_m\downarrow\theta_{m-1}}\{C(\mu^c(\theta_m,\theta_{m-1}),\theta_m)+\xi\beta(\theta_m)C_\theta(\mu^c(\theta_m,\theta_{m-1}),\theta_m)\} = C(\mu^c(\theta_{m-1},\theta_{m-1}),\theta_{m-1})+\xi\beta(\theta_{m-1})C_\theta(\mu^c(\theta_{m-1},\theta_{m-1}),\theta_{m-1})
$$

and by L'Hospital's rule the virtual capacity cost

$$
\lim_{\theta_m\downarrow\theta_{m-1}} VC(\mu^c(\theta_m,\theta_{m-1}),\theta_m,\theta_{m-1}) = C(\mu^c(\theta_{m-1},\theta_{m-1}),\theta_{m-1})+\xi\beta(\theta_{m-1})C_\theta(\mu^c(\theta_{m-1},\theta_{m-1}),\theta_{m-1}),
$$

$$
H(\theta_{m-1},\theta_{m-1}) = \lim_{\theta_m\downarrow\theta_{m-1}}\{UV^m - C(\mu^c(\theta_m,\theta_{m-1}),\theta_m)-\xi\beta(\theta_m)C_\theta(\mu^c(\theta_m,\theta_{m-1}),\theta_m)\},
$$

and therefore

$$
\begin{aligned}
H(\theta_{m-1},\theta_{m-1}) &= UV^m - C(\mu^c(\theta_{m-1},\theta_{m-1}),\theta_{m-1}) \\
&\quad -\xi\beta(\theta_{m-1})C_\theta(\mu^c(\theta_{m-1},\theta_{m-1}),\theta_{m-1}) \\
&> UV^{m+1} - C(\mu^c(\theta_{m+1},\theta_{m-1}),\theta_{m-1}) \\
&\quad -\xi\beta(\theta_{m-1})C_\theta(\mu^c(\theta_{m+1},\theta_{m-1}),\theta_{m-1}).
\end{aligned}
$$

Thus, $\frac{d}{d\theta_m}E\{NV;M\}>0$ when θ_m is in a neighborhood of θ_{m-1}. Again by L'Hospital's rule,

$$
\lim_{\theta_m\downarrow\theta_{m+1}} VC(\mu^c(\theta_{m+1},\theta_m),\theta_{m+1},\theta_m) = C(\mu^c(\theta_{m+1},\theta_{m+1}),\theta_{m+1})+\xi\beta(\theta_{m+1})C_\theta(\mu^c(\theta_{m+1},\theta_{m+1}),\theta_{m+1}).
$$

Hence

$$
\begin{aligned}
H(\theta_{m+1},\theta_{m+1}) &= \lim_{\theta_m \uparrow \theta_{m+1}} \{UV^{m+1} - C(\mu^c(\theta_{m+1},\theta_m),\theta_m) \\
&\quad -\xi\beta(\theta_m)C_\theta(\mu^c(\theta_{m+1},\theta_m),\theta_m)\} \\
&= UV^{m+1} - C(\mu^c(\theta_{m+1},\theta_{m+1}),\theta_{m+1}) \\
&\quad -\xi\beta(\theta_{m+1})C_\theta(\mu^c(\theta_{m+1},\theta_{m+1}),\theta_{m+1}) \\
&> UV^{m} - C(\mu^c(\theta_{m+1},\theta_{m-1}),\theta_{m+1}) \\
&\quad -\xi\beta(\theta_{m+1})C_\theta(\mu^c(\theta_{m+1},\theta_{m-1}),\theta_{m+1}),
\end{aligned}
$$

and thereby $\frac{d}{d\theta_m}E\{NV;M\} < 0$ when θ_m is in a neighborhood of θ_{m+1}. So (5.27) changes its sign from positive to negative at least once, and consequently the constraints (5.23) are non-binding and the optimal partition is never degenerate.

It now must be shown that the right hand side of (5.26) is single-peaked in θ_m, so the solution to (5.25) gives the unique optimal partition. By the negative definiteness of $H(\lambda,\mu,\theta_m,\theta_{m-1})$, it can be written as a function of μ, $H(\mu,\theta_m,\theta_{m-1})$, and $H(\mu,\theta_m,\theta_{m-1})$ is strictly concave in μ. Thus,

$$UV(\mu) - C(\mu,\theta_m) - \xi\beta(\theta_m)C_\theta(\mu,\theta_m)$$

is concave in μ as well. Since

$$
\begin{aligned}
&\max_{\mu}\ UV(\mu) - C(\mu,\theta_m) - \xi\beta(\theta_m)C_\theta(\mu,\theta_m) = \\
&\qquad UV(\mu^c(\theta_m,\theta_m)) - C(\mu^c(\theta_m,\theta_m),\theta_m) - \xi\beta(\theta_m)C_\theta(\mu^c(\theta_m,\theta_m),\theta_m)
\end{aligned}
$$

and $\mu^c(\theta_m,\theta_{m-1})$ is decreasing in both θ_m and θ_{m-1},

$$UV(\mu^c(\theta_m,\theta_{m-1})) - C(\mu^c(\theta_m,\theta_{m-1}),\theta_m) - \xi\beta(\theta_m)C_\theta(\mu^c(\theta_m,\theta_{m-1}),\theta_m)$$

is strictly decreasing in both θ_m and θ_{m-1}. Then

$$\sum_{j=m}^{m+1} H(\theta_m,\theta_{m-1})[F(\theta_m) - F(\theta_{m-1})]$$

must be single-peaked in θ_m, thus establishing the uniqueness of the optimal partition.||

Proposition 5.2 establishes the existence of a unique, non-degenerate optimal partition. This also suggests that the expected organizational net value is increasing in the number of intervals provided that the partitions are optimally constructed. Consequently, the optimal partition for any finite M can be obtained by solving the unconstrained maximization problem (5.22). Equation (5.25), an "arbitrage" condition, is a well-defined second-order nonlinear difference equation in θ_m with the initial and terminal conditions $\theta_0 = \underline{\theta}$ and $\theta_M = \overline{\theta}$, respectively. With arbitrary forms of functions and distribution, this set of second-order nonlinear difference equations is very difficult to solve even numerically. But the idea of solving the optimal partition is simple. Given the initial condition $\theta_0 = \underline{\theta}$, θ_2 can be determined uniquely if θ_1 is specified, since the right hand side of (5.25) is decreasing in θ_{m+1}. Once θ_2 has been determined, θ_3 can also be determined accordingly. Since θ_2 and thus θ_m for all $m \in \{3, \ldots, M\}$ are continuous and monotone in θ_1, the optimal partition is obtained once the θ_1 has been found with which the terminal condition, $\theta_M = \overline{\theta}$, is satisfied. I provide a simple numerical example in the following section.

5.6 Example 5.2

I continue to make use of the assumptions given in Example 5.1, but I now allow M to be greater than 2 and focus on the case where $\xi = 1$. I derive the results for two centralized mechanisms, first for a mechanism governed by an equally spaced partition, and then for a mechanism with the optimal partition.

The Equally Spaced Partition. A partition is equally spaced if, for a given M, Θ is partitioned into M equally spaced intervals, i.e., $\theta_m - \theta_{m-1} = \frac{\overline{\theta}-\underline{\theta}}{M} \equiv r(m)$, for all $m \in \{1, 2, \ldots, M\}$. For this example, it can easily be verified that (5.16) is negative definite with respect to λ and μ for each m. Then, from Lemma 5.2, the incentive compatible budget allocation is:

$$\begin{aligned} T(m) &= \theta_m \, \mu(\theta_m, \theta_{m-1}) + \sum_{j=m+1}^{M} \{[\theta_j - \theta_{j-1}] \, \mu(\theta_j, \theta_{j-1})\} \\ &= \theta_m \, \mu(\theta_m, \theta_{m-1}) + r(M) \sum_{j=m+1}^{M} \mu(\theta_j, \theta_{j-1}), \end{aligned}$$

and for any $\theta \in \Theta_m$, the excess budget allocation is:

$$(\theta_m - \theta)\,\mu(\theta_m, \theta_{m-1}) + r(M) \sum_{j=m+1}^{M} \mu(\theta_j, \theta_{j-1}).$$

Letting the superscript "e" denote this equally spaced case,

$$H^e(\lambda^e(\theta_m,\theta_{m-1}), \mu^e(\theta_m,\theta_{m-1}); \theta_m,\theta_{m-1}) = V(\lambda^e(\theta_m,\theta_{m-1}))$$
$$-\lambda^e(\theta_m,\theta_{m-1}) W(\lambda^e(\theta_m,\theta_{m-1}), \mu^e(\theta_m,\theta_{m-1})) - \gamma^e(\theta_m,\theta_{m-1})\mu^e(\theta_m,\theta_{m-1}),$$

where

$$\begin{aligned}
\lambda^e(\theta_m,\theta_{m-1}) &= \left[\frac{k - \sqrt{\gamma^e(\theta_m,\theta_{m-1})}}{\gamma^e(\theta_m,\theta_{m-1})}\right]^2; \\
\mu^e(\theta_m,\theta_{m-1}) &= \frac{k\left[k - \sqrt{\gamma^e(\theta_m,\theta_{m-1})}\right]}{\gamma^e(\theta_m,\theta_{m-1})^2}; \\
\gamma^e(\theta_m,\theta_{m-1}) &= \theta_m + (\theta_{m-1} - \underline{\theta}) \\
&= \theta_m + (m-1)\,r(M) \\
&= \underline{\theta} + (2m-1)\,r(M).
\end{aligned}$$

Clearly, $\mu^e(\theta_m,\theta_{m-1})$ is decreasing in m, so the mechanism is globally incentive compatible. Letting $H^e(\theta_m,\theta_{m-1}) \equiv H(\lambda^e, \mu^e; \theta_m,\theta_{m-1})$,

$$H^e(\theta_m,\theta_{m-1}) = \gamma^e(\theta_m,\theta_{m-1})\,\lambda^e(\theta_m,\theta_{m-1}),$$

and so the expected organizational net value with an equally spaced partition containing M intervals is:

$$E\{NV^e; M\} = \sum_{m=1}^{M} \frac{\left[k - \sqrt{\gamma^e(\theta_m,\theta_{m-1})}\right]^2}{M\gamma^e(\theta_m,\theta_{m-1})}.$$

The Optimal Partition. With $\xi = 1$, for any arbitrary partition,

$$H^c(\lambda^c, \mu^c; \theta_m,\theta_{m-1}) = V(\lambda^c) - \lambda^c W(\lambda^c,\mu^c) - \gamma^c(\theta_m,\theta_{m-1})\mu^c$$

where $\gamma^c(\theta_m,\theta_{m-1}) = \theta_m + (\theta_{m-1} - \underline{\theta})$, and

$$\begin{aligned}
\lambda^c(\theta_m,\theta_{m-1}) &= \left[\frac{k - \sqrt{\gamma^c(\theta_m,\theta_{m-1})}}{\gamma^c(\theta_m,\theta_{m-1})}\right]^2 \\
\mu^c(\theta_m,\theta_{m-1}) &= \frac{k\left[k - \sqrt{\gamma^c(\theta_m,\theta_{m-1})}\right]}{\gamma^c(\theta_m,\theta_{m-1})^2} \\
H^c(\theta_m,\theta_{m-1}) &= \gamma^c(\theta_m,\theta_{m-1})\lambda^c(\theta_m,\theta_{m-1})
\end{aligned}$$

Substituting $\gamma^c(\theta_m,\theta_{m-1})\lambda^c(\theta_m,\theta_{m-1})$ for $H^c(\theta_m,\theta_{m-1})$ in (5.27) gives the following set of first-order conditions:

$$\begin{aligned} 0 &= \gamma^c(\theta_m,\theta_{m-1})\lambda^c(\theta_m,\theta_{m-1}) - \gamma^c(\theta_{m+1},\theta_m)\lambda^c(\theta_{m+1},\theta_m) \\ &\quad -(\theta_{m+1}-\theta_m)\mu^c(\theta_{m+1},\theta_m) - (\theta_m - \theta_{m-1})\mu^c(\theta_m,\theta_{m-1}), \end{aligned}$$

for $m \in \{1,\ldots,M-1\}$. This set of first-order conditions cannot be solved explicitly even with the simple functions that I assume, so they must be solved numerically, as shown in Figure 5.2.

Discussion. Figure 5.2 depicts the optimal partition with M ranging from 1 to 10 for the case where $k = 5$ and $\Theta = [1,2]$. For this example, the lengths of the intervals contained in the optimal partition are increasing in m, i.e., $\theta_m - \theta_{m-1}$ is increasing in m. The equally spaced partition fails to be optimal even though the underlying support is uniformly distributed. Figures 5.3 and 5.4 respectively show the optimal capacity for the equally spaced case and for the optimal partition case. The convex curve corresponds to the capacity function, $\mu^*(\theta)$, when the communication is unlimited. Notice that for this example, regardless of how the partition is formed, $\mu^c(m)$ must intersect with $\mu^*(\theta)$ for all m. To see this, letting $\gamma^*(\theta)$ denote the virtual marginal cost when the communication is unlimited, then

$$\gamma^*(\theta) = \theta + (\theta - \underline{\theta}),$$

and therefore

$$\gamma^*(\theta_{m-1}) < \gamma^c(\theta_m,\theta_{m-1}) < \gamma^*(\theta_m).$$

Since $\mu^*(\theta)$ is monotone decreasing in θ, $\mu^c(m)$ must intersect $\mu^*(\theta)$ for all m.

In Figure 5.5, I compare the net value that can be generated by the optimal partition with that generated by the equally spaced partition. For this particular example, an equally spaced partition can provide almost the same performance as the optimal partition can, especially when M is large.

As discussed earlier, when $M > 1$, the expected organizational net value need not be monotone in M unless all the partitions are optimally constructed. It is easy to construct an example such that $E\{NV;2\} > E\{NV;3\}$. For instance, using the same example with an equally spaced partition, $E\{NV;2\} = 7.0886$. But if there is a partition with $\theta_1 = 1.9$ and $\theta_2 = 1.95$, $E\{NV;3\} = 6.5879$. Thus the way the partition is formed will affect the IS department's informational rent, and thus the expected organizational net value.

If the message set and the corresponding partition can be affected by the central management, e.g., investing time and effort to understand more about information technology and its IS department's operating environment, then a "design" issue arises. For example, when $\mathcal{M} = \{\text{high, low}\}$, the central management may want to spend more time on IS operations in order to be able to further partition the lower interval corresponding to the message "low" into two intervals corresponding to the messages "medium" and "low." However, due to the complexity of the problem, I am not able to provide a general characterization.

When the total budget allocation is increasing in the IS department's efficiency (e.g., $\gamma\mu$ is decreasing in m in our example), the scale of the organization's information processing operations (the effective capacity) as well as the total budget allocated to the IS department may be larger for a more efficient IS department. It is not unreasonable, then, for the central management to exert more effort to evaluate the IS department's budget proposal when the IS department is efficient and a large scale operation is desired. For my abstract model of limited communication, this may correspond to a more dense partitioning at the lower end of the parameter space, Θ, as suggested by the example.

Furthermore, if the central management can choose a message set of an arbitrary number of messages with some costs, $\Psi(M)$ can be defined as the cost associated with the communication system $COM(M)$ that is capable of partitioning Θ into an arbitrary partition of M intervals. $\Psi(M)$ is assumed to be increasing in M with $\Psi(1) = 0$ and $\Psi(\infty) = \infty$. Then $COM(1)$ corresponds to a "null" message set and $COM(\infty)$ corresponds to unlimited communication. Or, alternatively, a "bit" can represent the unit of the communication budget so that, given the number of the bits, b, $\Psi(b)$ is the cost of the communication system and $M(b) = 2^b$ is the number of messages that can be supported by b bits. If the number of messages is considered directly, then, letting $E\{NV^c; M\}$ denote the optimal expected organizational net value with the optimal communication system and I be the set of positive integers, the central management's optimal choice of a communication system is the solution to the following problem:

$$\max_{M \in I} E\{NV^c; M\} - \Psi(M).$$

Since $E\{NV^c; M\}$ is monotone in M bounded from above by $E\{NV^c; \infty\}$, and bounded from below by $E\{NV^c; 1\}$, the above program has a solution, provided that $\Psi(M)$ increases fast enough. Since $E\{NV^c; M\}$ is increasing in M, if it is also concave and $\Psi(M)$ is (weakly) convex, then the solution can be characterized by comparing the marginal expected organizational net value with the marginal cost of the communication system. That is, the largest M must be found such that $E\{NV^c; M\} - E\{NV^c; M-1\} \geq \Psi(M) - \Psi(M-1)$ and $E\{NV^c; M+1\} - E\{NV^c; M\} < \Psi(M+1) - \Psi(M)$. For example, from Figure 5.5, it is clear that $E\{NV^c; M\}$ is concave in M. Assuming that $\Psi(M) = \omega M$, then the optimal communication system $COM(M^c)$ is obtained by choosing a system such that $E\{NV^c; M^c\} - E\{NV^c; M^c - 1\} \geq \omega$ and $E\{NV^c; M^c + 1\} - E\{NV^c; M^c\} < \omega$. For instance, if $\omega = 0.1$, it can be easily verified that $M^c = 3$.

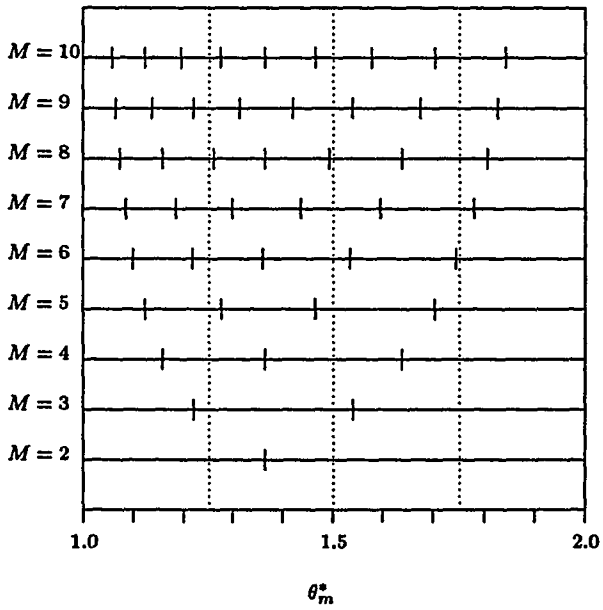
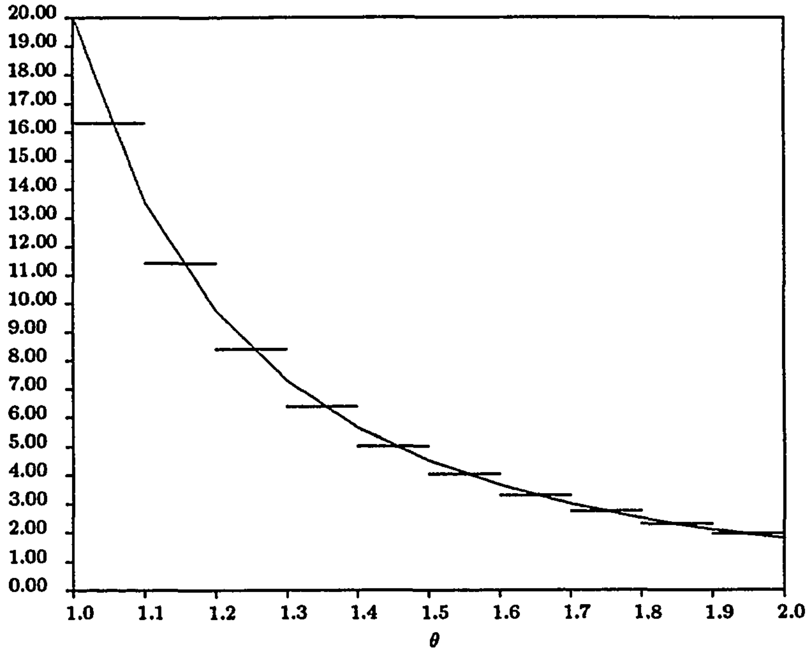
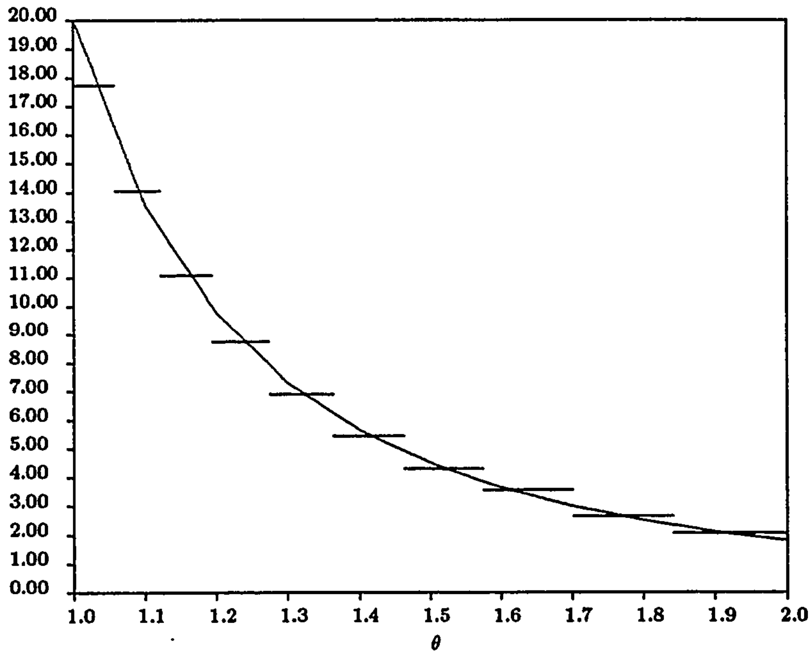


FIGURE 5.2: EXAMPLE 5.1—THE OPTIMAL PARTITION AS M VARIES FROM 2 TO 10.



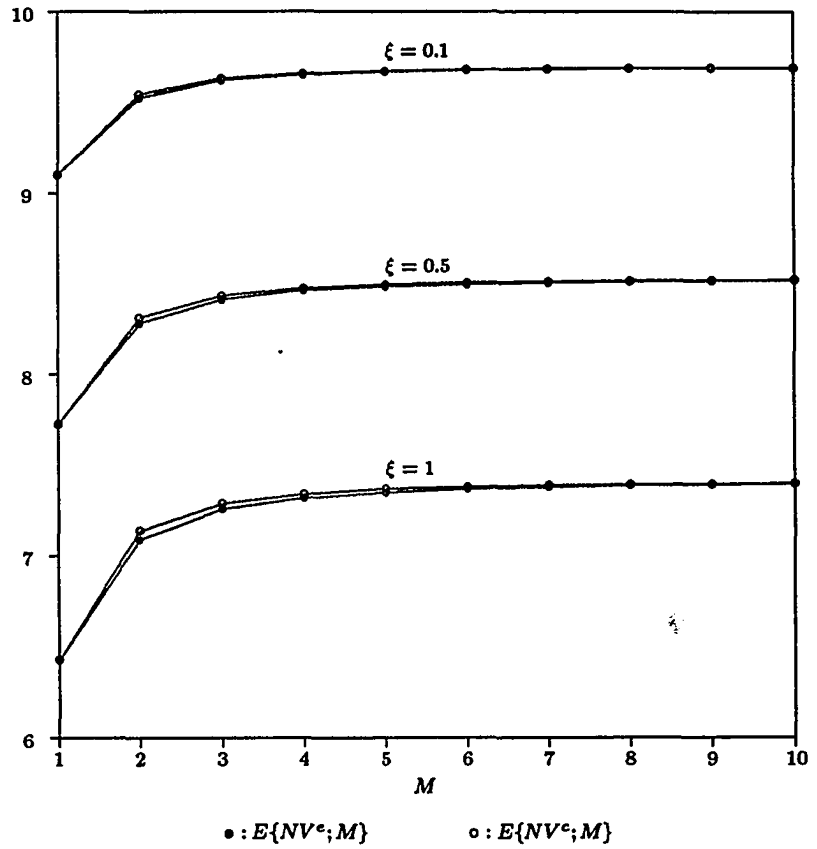
The capacity functions $\mu^*(\theta)$ for the unlimited communication case and $\mu^e(m)$ for the limited communication case with an equally spaced partition of 10 intervals. The convex curve depicts $\mu^*(\theta)$ whereas the step function is $\mu(m)$.

FIGURE 5.3: EXAMPLE 5.2—CASE WITH THE EQUALLY SPACED PARTITION.



The capacity functions $\mu^*(\theta)$ for the unlimited communication case and $\mu(m)$ for the limited communication case with the optimal partition of 10 intervals. The convex curve depicts $\mu^*(\theta)$ whereas the step function is $\mu^c(m)$.

FIGURE 5.4: EXAMPLE 5.2—CASE WITH THE OPTIMAL PARTITION.



The expected organizational net values with an equally spaced partition ($E\{NV^e; M\}$) and with the optimal partition ($E\{NV^c; M\}$), as M varies from 1 to 10 and ξ varies from 0.1 to 1.

FIGURE 5.5: EXAMPLE 5.2—THE COMPARISON BETWEEN THE EQUALLY SPACED AND THE OPTIMAL PARTITIONS.

Chapter 6

Imperfect, Noiseless Monitoring

6.1 Introduction

In most existing literature studying centralized organizational decision-making mechanisms, the range over which an economic agent can misrepresent private information is typically bounded only by the mechanism designer's prior beliefs. This assumption is relaxed in this chapter as I study the central management's ability to control its IS department with asymmetric information. I investigate the central management's ability to control its IS department in an environment where the central management possesses an imperfect but noiseless monitoring (partially verifiable) information system. In particular, I want to know whether it is always optimal for the central management to elicit the IS manager's truth-revelation when her report is partially verifiable.

Traditional principal-agent models dealing with moral hazard focus on the observability of the agent's action through monitoring (e.g., Dye [29]; Holmstrom [52]; Singh [101]). In this research, the monitoring technology typically is modeled as a sampling process capable of generating a signal correlated with the agent's action. The monitoring policy, then, is characterized as the probability of audit. Baron and Besanko [8] study similar types of monitoring policies in a regulatory setting without moral hazard. In these models, the realized cost is a random variable that cannot be completely controlled by the agent, whereas, in my model, the IS department has complete control over the realized cost by excessive investment. Consequently, the auditing policies that Baron and Besanko study are irrelevant in our environment. My investigation of the IS-related decision rules with different monitoring systems is based on the results established by Guesnerie and Laffont [37]. The motivation to study this type of monitoring is grounded in "the observation that although lying is a problem within most organizations, there is only a limited range of distortion against which the system must guard itself" (Guesnerie and Laffont [37], p. 448). For tractability, I assume that the communication is unlimited, $\mathcal{M} = \Theta$, and that the set of feasible systems is continuous, $\mathcal{K} = \mathcal{R}_+$.

Monitoring is the central management's ability to detect the IS manager's false report. That is, my focus is on the usefulness of monitoring systems in verifying the IS manager's report. As in Guesnerie and Laffont [37], I focus on cases where monitoring technology is imperfect but noiseless. A monitoring system is imperfect when the central management cannot detect and verify the IS manager's false report as long as it is not "too big" a lie. A monitoring system is noiseless if the central management can detect and verify the IS manager's false report with certainty whenever the difference between the false report and the truth is beyond a certain limit specified by the monitoring system. Formally, let $\mathcal{N} : \Theta \rightarrow \Theta$ be the correspondence determining the admissible reports. Then, for each θ, $\mathcal{N}(\theta) \subseteq \Theta$ is the set of reports to which the IS manager's report is restricted. Of course, in order for the IS manager's truth-revelation to be always possible, I assume that $\theta \in \mathcal{N}(\theta)$ for all $\theta \in \Theta$. Like Guesnerie and Laffont [37], I assume that the central management does not have an independent observation of θ. Rather, the central management can only observe a binary variable generated by the monitoring system and the value of this variable is (non-stochastically or noiselessly) jointly determined by the true θ and the IS manager's report, $\hat{\theta}$. The value of this variable indicates whether or not $\hat{\theta} \in \mathcal{N}(\theta)$. Notice that, if $\mathcal{N}(\theta)$ varies with θ, it is important not to allow the central management to observe the entire set $\mathcal{N}(\theta)$; otherwise, the central management would be able to infer the true θ itself.

When the IS manager's utility can be constructed to be very bad whenever the central management learns that $\hat{\theta} \notin \mathcal{N}(\theta)$, the IS manager will never send such a report. So the proper interpretation of the monitoring system is that the central management also has some information, and that the central management can act on this information to inflict severe punishment on the IS manager in some circumstances.

Two types of monitoring technology are studied: the *parameter-bound* and the *space-partition* monitoring systems. A parameter-bound system imposes a bound on the information that the IS manager can misrepresent. If the IS manager misrepresents her information but within the bound, she can be sure her misrepresentation will go undetected; otherwise, her misrepresentation will be detected with certainty. Monitoring systems of this type fail to satisfy a so-called "nested range condition," which is necessary for the revelation principle to be valid (Green and Laffont [37]). That is, the outcomes attainable by a mechanism that does not demand truthful revelations are not necessarily attainable by a truth-revelation mechanism. For the problem that I study, monitoring systems of this type are of no value to the central management in inducing truth-revelation. Thus, if the central management possesses a monitoring system of this type and the parameter-bound is sufficiently narrow, the central management may be better off by simply treating the IS manager's report as truth. The effect of this type of monitoring technology and the optimal design for such monitoring systems under the non-truth-revelation mechanism will be analyzed in Section 6.2.

The space-partition monitoring technology partitions Θ into intervals so that the IS manager cannot misrepresent her information across intervals. Since systems of this type satisfy the "nested range condition," the revelation principle remains intact. This type of monitoring system allows the central management to treat each interval independently when designing the mechanism. As a result, the globally optimal mechanism consists of a set of decision rules, one for each interval. I further derive the optimality condition for designing the optimal monitoring system with a fixed number intervals that the system is capable of partitioning. In Section 6.3, I show that, as in the limited communication case, when the systems are costless, a system with a larger message set does not necessarily generate a higher organizational net value unless the partition is finer or optimally constructed.

Although my models of imperfect monitoring technologies are abstract, they can nevertheless serve as a first-order approximation of real-world monitoring technologies.

For instance, an imperfect monitoring system can be considered a model of the central management's attempt to identify the efficiency of its IS operation by comparing its level of investment and the scale and the scope of IS operations with those of other companies in the same industry. This kind of information usually is public and not very costly to obtain. The results of the comparisons may give the central management a rough idea about whether or not the performance of its IS operations is reasonable. Another example arises when the central management hires an external consulting firm to evaluate its IS department.[1] Even without these extra monitors, misuses of organizational resources by a large amount in general is more difficult for the IS manager to hide. In terms of a budget proposal, if the proposed budget allocation is too "unreasonable" given the state of the organization's operating environment, the IS manager may well have a hard time giving a credible explanation for what she is proposing. It is reasonable to assume, therefore, that the IS manager can misrepresent her information and go undetected only if she does not exaggerate too much. Of course, these monitoring methods cannot perfectly identify all possible misuses of an organization's resources and typically can only give a very rough estimate.

Since the monitoring systems have no effect when the IS department is organized as a profit center, I do not discuss the profit center case in this chapter. Before studying the effects of these monitoring technologies in detail, it is helpful to first introduce the "nested range condition" provided in Green and Laffont [37]:

Nested Range Condition. A monitoring technology satisfies the "nested range condition" provided that, for any three distinct elements $\theta_1, \theta_2, \theta_3 \in \Theta$, if $\theta_2 \in \mathcal{N}(\theta_1)$ and $\theta_3 \in \mathcal{N}(\theta_2)$, then $\theta_3 \in \mathcal{N}(\theta_1)$.

6.2 Parameter-Bound Monitoring Systems

A parameter-bound monitoring system imposes a bound on the IS manager's ability to misrepresent her private information without being detected. Let $MON(\delta)$ denote a

[1] Delegating monitoring may introduce another layer of agency issues (Olivella [89]). These issues will not be pursued here.

monitoring system with bound δ. Then, given the IS department's true cost parameter θ, the IS manager can over or understate her department's cost parameter within the range $[\theta - \delta, \theta + \delta]$ without being detected. Since, as shown in Chapter 3, the IS manager tends to exaggerate her department's costs, I restrict my attention without loss of generality to the range $[\theta, \theta + \delta]$, where θ is the true cost parameter. Thus, given a monitoring system $MON(\delta)$, $\mathcal{N}(\theta) = [\theta, \theta + \delta]$ and moves with the true cost parameter. If the IS manager's report $\hat{\theta}(\theta) \in \mathcal{N}(\theta)$, her misrepresentation will go undetected; if her report $\hat{\theta}(\theta) \notin \mathcal{N}(\theta)$, her misrepresentation will be detected with certainty. To see that the parameter-bound technology does not satisfy the "nested range condition" as long as $\delta > 0$, let $\theta_2 \in (\theta_1, \theta_1 + \delta]$; then any $\theta_3 \in (\theta_1 + \delta, \theta_2 + \delta]$ is not in $\mathcal{N}(\theta_1)$.

With this type of monitoring technology, the central management can still induce the IS manager's truth-revelation by an appropriately designed mechanism, but there may exist a non-truth-revelation mechanism that can outperform it. This conclusion is provided as Theorem 1 in Green and Laffont [37]. The following proposition shows that, if the central management insists on implementing a truth-revelation mechanism, the parameter-bound monitoring technology is of no value for the central management in inducing the IS manager's truth-revelation as long as $\delta > 0$; i.e., unless a monitoring system of this type is perfect, it fails to reduce the informational rent that the IS manager can command under a truth-revelation mechanism.

PROPOSITION 6.1 *A monitoring system $MON(\delta)$ has no value for implementing a truth-revelation mechanism as long as $\delta > 0$.*

PROOF. It is sufficient to show that the excess budget allocation required for inducing the IS manager's truth-revelation when the central management possesses a monitoring system $MON(\delta)$, $\delta > 0$ is the same as in the case where the central management does not possess such a system.

Let $T(\hat{\theta})$ be the budget allocated to the IS department if the manager reports $\hat{\theta}$. If the true cost parameter is θ, incentive compatibility requires:

$$T'(\hat{\theta}) - C_\mu(\mu(\hat{\theta}), \theta)\Big|_{\hat{\theta}=\theta} \equiv 0, \tag{6.1}$$

implying

$$S'(\theta) = -C_\theta(\mu(\theta), \theta), \tag{6.2}$$

where

$$S(\theta) = T(\theta) - C(\mu(\theta), \theta).$$

Since the IS manager can only exaggerate up to $\theta + \delta$,

$$S(\theta) = S(\theta + \delta) + \int_\theta^{\theta+\delta} C_\theta(\mu(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta} \tag{6.3}$$

by integrating (6.2), where

$$S(\theta + \delta) = T(\theta + \delta) - C(\mu(\theta + \delta), \theta + \delta).$$

But, in order to induce the IS manager's truth-revelation when the realized cost parameter is $\theta + \delta$,

$$S(\theta + \delta) = S(\theta + 2\delta) + \int_{\theta+\delta}^{\theta+2\delta} C_\theta(\mu(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta}. \tag{6.4}$$

Thus,

$$S(\theta) = S(\theta + 2\delta) + \int_\theta^{\theta+2\delta} C_\theta(\mu(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta}. \tag{6.5}$$

By induction,

$$S(\theta) = S(\theta + n\delta) + \int_\theta^{\theta+n\delta} C_\theta(\mu(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta}, \tag{6.6}$$

where $n = \lfloor \frac{\overline{\theta}-\theta}{\delta} \rfloor$, the largest integer satisfying $\theta + n\delta \leq \overline{\theta}$. For $\theta \in [\overline{\theta} - \delta, \overline{\theta}]$, incentive compatibility requires:

$$S(\theta) = S(\overline{\theta}) + \int_\theta^{\overline{\theta}} C_\theta(\mu(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta}, \tag{6.7}$$

where $S(\overline{\theta}) \geq 0$. Combining (6.6) and (6.7) yields:

$$S(\theta) = S(\overline{\theta}) + \int_\theta^{\overline{\theta}} C_\theta(\mu(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta}, S(\overline{\theta}) \geq 0, \tag{6.8}$$

which is exactly the same as required for the case where the central management does not possess a monitoring system. ||

Proposition 6.1 shows that the parameter-bound monitoring technology has no value in implementing a truth-revelation mechanism unless $\delta = 0$ (perfect monitoring). At first glance, this result seems surprising. However, it is intuitively clear that the central management must induce the IS manager to report truthfully for all possible realizations of the cost parameter ex ante if it insists on implementing a truth-revelation mechanism. Then the budget allocation for $\theta + \delta$ cannot be arbitrary or simply set to equal zero if the mechanism must induce the IS manager's truth-revelation.

If the bound δ changes its size so that the admissible reports form an order space on Θ, i.e., for all $\theta \in \Theta$, $\mathcal{N}(\theta) = [\theta, \overline{\theta}]$, then it is clear that $\mathcal{N}(\theta)$ satisfies the nested range condition. This monitoring technology can prevent the IS manager from understating the costs. But, since the IS manager only has incentives to overstate the costs, monitoring systems of this type are useless.

Naive Mechanisms with Parameter-bound Monitoring

Although a parameter-bound monitoring system is of no value for implementing a truth-revelation mechanism, it may be valuable to a naive central management as long as it in effect imposes a bound on the information that the IS manager can misrepresent. When the central management is naive, then given θ, the IS manager determines her report by solving:

$$\max_{\hat{\theta}} S^n(\hat{\theta}; \theta) \tag{6.9}$$

subject to

$$\hat{\theta} \leq \min\{\theta + \delta, \overline{\theta}\}, \qquad \forall \theta \in \Theta, \tag{6.10}$$

where $S^n(\hat{\theta}; \theta) \stackrel{\text{def}}{=} C(\mu^n(\hat{\theta}), \hat{\theta}) - C(\mu^n(\hat{\theta}), \theta)$ and $\mu^n(\hat{\theta})$ is the solution to the naive central management's problem when the report is $\hat{\theta}$, i.e., the solution to the following program:

$$NV^n(\hat{\theta}) = \max_{\lambda, \mu} V(\lambda) - \lambda W(\lambda, \mu) - C(\mu, \hat{\theta}).$$

Since

$$S^n_{\hat{\theta}}(\hat{\theta}, \theta)\Big|_{\hat{\theta}=\theta} = C_\theta(\mu^n(\theta), \theta) > 0,$$

the IS manager will exaggerate her department's cost parameter when the central management is naive. Let $\theta^n(\theta)$ be the solution to the program (6.9)–(6.10). Assuming that the bound δ is sufficiently narrow, so that the constraint (6.10) is binding for all $\theta \in \Theta$, then the IS manager's report is $\theta^n(\theta) = \theta + \delta$ for all $\theta \in [\underline{\theta}, \overline{\theta} - \delta)$ and $\theta^n(\theta) = \overline{\theta}$ for all $\theta \in [\overline{\theta} - \delta, \overline{\theta}]$. Consequently, when the organization is controlled by a naive central management with a monitoring system $MON(\delta)$, the expected organizational net value is:

$$E\{NV^n(\theta);\delta\} = \int_{\underline{\theta}}^{\overline{\theta}-\delta} NV^n(\theta+\delta)\,dF(\theta) + \int_{\overline{\theta}-\delta}^{\overline{\theta}} NV^n(\overline{\theta})\,dF(\theta).$$

Obviously, $E\{NV^n(\theta);\delta\}$ approaches the full information case as $\delta \to 0$. The following theorem characterizes the behavior of the expected organizational net value with respect to the variation of δ.

PROPOSITION 6.2 *If the constraint (6.10) is binding, the expected organizational net value $E\{NV^n(\theta);\delta\}$ is decreasing in δ. Furthermore, if $C_{\theta\theta}(\mu,\theta) \geq 0$, $E\{NV^n(\theta);\delta\}$ is convex in δ.*

PROOF. Suppose (6.10) is binding; differentiating $E\{NV^n;\delta\}$ with respect to δ yields:

$$\begin{aligned}
\frac{d}{d\delta}E\{NV^n(\theta);\delta\} &= -NV^n(\overline{\theta})f(\overline{\theta}-\delta) + \int_{\underline{\theta}}^{\overline{\theta}-\delta} \frac{dNV^n(\theta+\delta)}{d\delta}\,dF(\theta) + NV^n(\overline{\theta})f(\overline{\theta}-\delta) \\
&= \int_{\underline{\theta}}^{\overline{\theta}-\delta} \frac{dNV^n(\theta+\delta)}{d\delta}\,dF(\theta) \\
&= -\int_{\underline{\theta}}^{\overline{\theta}-\delta} \left.\frac{\partial C(\mu^n(\theta+\delta),y)}{\partial y}\right|_{y=\theta+\delta} dF(\theta) \\
&< 0,
\end{aligned}$$

where the last equality follows from the envelope theorem. Hence the expected organizational net value $E\{NV^n(\theta);\delta\}$ is decreasing in δ when the constraint (6.10) is binding. Furthermore,

$$\begin{aligned}
\frac{d^2}{d\delta^2}E\{NV^n(\theta);\delta\} &= f(\overline{\theta}-\delta)\left.\frac{\partial C(\mu^n(\overline{\theta}),y)}{\partial y}\right|_{y=\overline{\theta}} \\
&\quad - \int_{\underline{\theta}}^{\overline{\theta}-\delta} \left\{\frac{\partial^2 C(x,y)}{\partial x\partial y}\frac{d\mu^n(y)}{dy} + \frac{\partial^2 C(x,y)}{\partial y\partial y}\right\}\Bigg|_{\substack{x=\mu^n(\theta+\delta)\\ y=\theta+\delta}} dF(\theta).
\end{aligned}$$

By the assumption that $C_\theta > 0$ and $C_{\theta\mu} > 0$ and by the fact that $\mu^n(\hat{\theta})$ is decreasing, the above expression is positive if $C_{\theta\theta} > 0$. ||

In fact, the proof reveals that $E\{NV^n(\theta);\delta\}$ is convex in δ as long as $C_{\theta\theta}$ is not too large a negative number. The second part of the theorem is merely a sufficient condition for $E\{NV^n(\theta);\delta\}$ to be convex in δ; in particular, when $C_{\theta\theta} = 0$, $\frac{d^2}{d\delta^2}E\{NV^n(\theta);\delta\} > 0$.

Consequently, the expected organizational net value is increasing and convex as the bound becomes tighter.

It can easily be shown that the decision rule $\mu^n(\cdot)$ under a naive mechanism with the parameter-bound monitoring technology cannot be truthfully implemented. Assume that $\tilde{\mu}(\cdot)$ truthfully implements $\mu^n(\cdot)$, i.e., $\tilde{\mu}(\hat{\theta}(\theta)) = \mu^n(\theta)$ and $\hat{\theta}(\theta) = \theta$. Given $\theta_1 \in \Theta$ and $\theta_2 \in \mathcal{N}(\theta_1)$,

$$\begin{aligned} \tilde{\mu}(\theta_1) &= \mu^n(\theta_1) \\ \tilde{\mu}(\theta_2) &= \mu^n(\theta_2). \end{aligned}$$

In response to $\tilde{\mu}(\cdot)$, $\hat{\theta}(\theta_1) = \theta_2$, since $S^n(\mu^n(\theta_1);\theta_1) < S^n(\mu^n(\theta_2);\theta_1)$ and $\theta_2 \in \mathcal{N}(\theta_1)$, where $S^n(\mu^n(\hat{\theta});\theta) = C(\mu^n(\hat{\theta}),\hat{\theta}) - C(\mu^n(\hat{\theta}),\theta)$. Thus, $\tilde{\mu}(\hat{\theta}(\theta_1)) = \mu^n(\theta_2) \neq \mu^n(\theta_1)$, and $\mu^n(\cdot)$ is not truthfully implementable.

Example 6.1: Naive Mechanisms. I again make use of Example 3.1. As has been shown in Chapter 3, if $3\underline{\theta} - \overline{\theta} > 0$ and

$$\frac{3}{8k} < \frac{3\theta - \overline{\theta}}{\sqrt{\overline{\theta}}\,(5\theta - \overline{\theta})},$$

$S(\hat{\theta};\theta)$ is concave in $\hat{\theta}$. Then the IS manager's optimal reporting strategy without the presence of a parameter-bound monitoring system is

$$\theta^n(\theta) = \begin{cases} \overline{\theta} & \text{if } \theta \geq \theta^o \\ \hat{\theta} & \text{where } g(\hat{\theta}) = \theta \text{ if } \theta \in [\underline{\theta}, \theta^o), \end{cases}$$

where θ^o is the solution to the equation $g(\overline{\theta}) = \theta^o$ and

$$g(\hat{\theta}) = \hat{\theta}\left[1 - \frac{2k - 2\sqrt{\hat{\theta}}}{4k - 3\sqrt{\hat{\theta}}}\right].$$

Lemma 3.2 shows that the IS manager tends to exaggerate the cost parameter more when the realized cost parameter is high. Thus, if $\theta^o < \overline{\theta}$ and $\delta < \hat{\theta}(\theta^o) - \theta^o$, there exist some states smaller than θ^o such that δ in effect imposes a bound, and therefore the $MON(\delta)$ system can generate positive value to the organization. Moreover, if $\delta \leq \hat{\theta}(\underline{\theta}) - \underline{\theta}$, the constraint (6.10) is binding for all θ. When this is the case,

$$\hat{\theta}(\theta) = \begin{cases} \overline{\theta} & \text{if } \theta \geq \theta^o \\ \theta + \delta & \text{otherwise.} \end{cases}$$

Consequently, the expected organizational net value

$$\begin{aligned} E\{NV^n(\theta);\delta\} &= \int_{\underline{\theta}}^{\overline{\theta}-\delta} NV^n(\theta+\delta)\,dF(\theta) + NV^n(\overline{\theta})\left(1-F(\overline{\theta}-\delta)\right) \\ &= \int_{\underline{\theta}}^{\overline{\theta}-\delta} \frac{(k-\sqrt{\theta+\delta})^2}{\theta+\delta}\,dF(\theta) + \frac{(k-\sqrt{\overline{\theta}})^2}{\overline{\theta}}\left(1-F(\overline{\theta}-\delta)\right) \\ &= \left[\frac{k^2\ln(\theta+\delta)-4k\sqrt{\theta+\delta}+\theta}{\overline{\theta}-\underline{\theta}}\right]_{\underline{\theta}}^{\overline{\theta}-\delta} \\ &\quad + \frac{(k-\sqrt{\overline{\theta}})^2}{\overline{\theta}}\left(1-F(\overline{\theta}-\delta)\right). \end{aligned} \tag{6.11}$$

For the specific example where $k = 5$, $\xi = 0.5$, and θ is distributed uniformly over $[1,2]$, $\hat{\theta}(\underline{\theta}) \approx 1.85$. Then if $\delta < 0.85$, the expected organizational net value is given by (6.11). Figure 6.1 depicts the expected organizational net value as δ decreases from 0.9 to 0. Clearly, when δ is small, it may be better for the central management to rely on monitoring and "play naive" instead of spending time and effort to design a truth-revelation mechanism.

Discussion. Since a parameter-bound monitoring system is of no value to an organization when it is implementing a truth-revelation mechanism, there is no need for the central management to implement a truth-revelation mechanism while exerting the effort to develop such a monitoring system. As shown in the previous example, when the bound δ is not very small (e.g., greater than 0.2), the truth-revelation mechanism can generate a higher organizational net value than the naive mechanism with a monitoring system can. However, when δ is small enough, the naive mechanism generates a larger expected organizational net value.

Although from the organization's standpoint it is best to have a costless perfect monitoring system so that the full information optimization can always be achieved, in real-world business organizations it is usually too costly to achieve perfect monitoring— if it is not impossible. The costs associated with a monitoring system may include the time and effort from the central management to design an appropriate reporting system and acquire expertise needed to evaluate local managers. Thus, when considering the ways to control the IS department, the central management may need to choose between the truth-revelation mechanism without monitoring and the naive mechanism with a costly monitoring system.

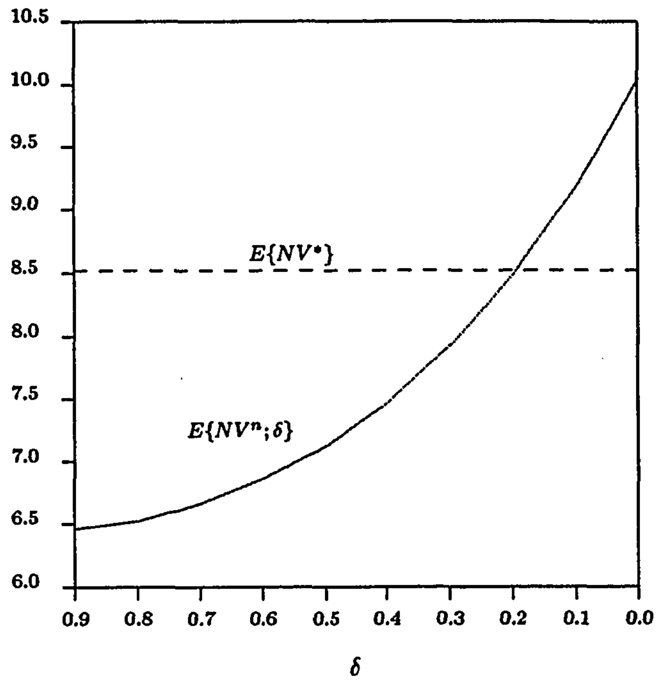
Typically the central management and the system users are not totally naive about information technologies and the organization's operation environment; the headquarters of the organization may well have a staff that is familiar with information technologies and is competent in evaluating the appropriateness of the IS budget requests. When this is the case, it may not be too costly for the central management to monitor the IS department effectively, i.e., to have a very small δ in the model. Thus when the central management's ability to monitor the IS department is very strong, it may be better off relying upon monitoring as a deterrent instead of exerting the effort to design a complex truth-revelation mechanism. But when the central management lacks the level of expertise required to monitor the IS department effectively, it may need to hire additional personnel or contract an outside consulting firm to oversee the IS department. In this case, an effective monitoring system may be too costly.

6.3 Space-Partition Monitoring Systems

A space-partition monitoring system partitions Θ into intervals. Let $MON(s)$ denote a monitoring system that partitions Θ into s intervals. Given a monitoring system $MON(s)$, let $\Theta_1 = [\theta_0, \theta_1]$, and $\Theta_i = (\theta_{i-1}, \theta_i], i \in \{2, \ldots, s\}$, where $\theta_0 = \underline{\theta}$ and $\theta_s = \overline{\theta}$. That is, $\Theta_i = \mathcal{N}(\theta)$, for all $\theta \in [\theta_{i-1}, \theta_i]$. So the range over which the IS manager can misrepresent her private information does not vary with the true cost parameter for all the parameter values within the same interval.

To see that the space-partition technology satisfies the "nested range condition," let $\theta_2 \in \mathcal{N}(\theta_1)$; then, since $\mathcal{N}(\theta_2) = \mathcal{N}(\theta_1)$, any $\theta_3 \in \mathcal{N}(\theta_2)$ must also be contained in $\mathcal{N}(\theta_1)$. Consequently, I can restrict my attention to truth-revelation mechanisms without loss of generality. Obviously, given this type of monitoring technology, the central management only has to worry about the truth-revelation by the IS manager within each interval. Thus, given that a realized $\theta \in \Theta_i$, incentive compatibility requires:

$$S_{\hat{\theta}}(\hat{\theta}; \theta)\Big|_{\hat{\theta}=\theta} \equiv 0,$$



The comparison between the expected organizational net values under the truth-revelation mechanism $E\{NV^*\}$ and the naive mechanism $E\{NV^n;\delta\}$ with the parameter-bound monitoring system as a function of the bound δ.

FIGURE 6.1: EXAMPLE 6.1—THE PARAMETER-BOUND MONITORING SYSTEM.

where $S(\hat{\theta};\theta) = T(\hat{\theta}) - C(\mu(\hat{\theta}),\theta)$. This implies

$$S'(\theta) = -C_\theta(\mu(\theta),\theta).$$

By integration, the excess budget allocation required for inducing the IS manager's truth-revelation is:

$$S(\theta) = S(\overline{\theta}) + \int_\theta^{\theta_i} C_\theta(\mu(\tilde{\theta}),\tilde{\theta})\,d\tilde{\theta}, \qquad \forall\theta \in \Theta_i,$$

where $S(\overline{\theta}) \geq 0$. Since it is always optimal for the central management to set $S(\overline{\theta}) = 0$,

$$S(\theta) = \int_\theta^{\theta_i} C_\theta(\mu(\tilde{\theta}),\tilde{\theta})\,d\tilde{\theta}, \qquad \forall\theta \in \Theta_i. \tag{6.12}$$

Since the central management's objective function is:

$$\sum_{i=1}^{s} \max_{\lambda(\cdot),\mu(\cdot)} \int_{\theta_{i-1}}^{\theta_i} \{V(\lambda(\theta)) - \lambda(\theta)W(\lambda(\theta),\mu(\theta)) - C(\mu(\theta),\theta) - \xi S(\theta)\}\,dF(\theta),$$

using (6.12) and integrating by parts yields:

$$\int_{\theta_{i-1}}^{\theta_i}\int_\theta^{\theta_i} C_\theta(\mu(\tilde{\theta}),\tilde{\theta})\,d\tilde{\theta}\,dF(\theta) = \int_{\theta_{i-1}}^{\theta_i} C_\theta(\mu(\theta),\theta)(F(\theta) - F(\theta_{i-1}))\,d\theta.$$

Then the central management's problem becomes:

$$\sum_{i=1}^{s} \max_{\lambda(\cdot),\mu(\cdot)} \int_{\theta_{i-1}}^{\theta_i} H(\lambda(\theta,\theta_{i-1}),\mu(\theta,\theta_{i-1}),\theta,\theta_{i-1})\,dF(\theta) \tag{6.13}$$

subject to

$$\mu(\theta,\theta_{i-1}) \text{ is non-increasing in } \theta \in \Theta_i, \qquad \forall i \in \{1,\ldots,s\}, \tag{6.14}$$

where

$$\begin{aligned} &H(\lambda(\theta,\theta_{i-1}),\mu(\theta,\theta_{i-1}),\theta,\theta_{i-1}) = \\ &\quad V(\lambda(\theta,\theta_{i-1})) - \lambda(\theta,\theta_{i-1})W(\lambda(\theta,\theta_{i-1}),\mu(\theta,\theta_{i-1})) \\ &\quad\quad -C(\mu(\theta,\theta_{i-1}),\theta) - \xi\alpha(\theta,\theta_{i-1})C_\theta(\mu(\theta,\theta_{i-1}),\theta), \end{aligned}$$

and

$$\alpha(\theta,\theta_{i-1}) \stackrel{\text{def}}{=} \frac{F(\theta) - F(\theta_{i-1})}{f(\theta)}.$$

Note that this program is just a set of subproblems in which each subproblem for a given interval is identical to the global problem when there is no monitoring, except that the support of the cost parameter is narrower. So it is clear that, if the capacity derived under the optimal mechanism when there is no monitoring is non-increasing, the optimal $\mu(\theta, \theta_{i-1})$ must also be non-increasing in θ within each Θ_i. The optimal capacity is not, however, globally non-increasing. But given the monitoring technology under consideration, this non-monotonicity will not destroy the incentive compatibility of the mechanism.

Assuming that the optimal capacity is non-increasing when there is no monitoring, then for each interval Θ_i, pointwise maximization with respect to λ and μ yields the following first-order conditions:

$$
\begin{aligned}
0 &= V'(\lambda) - W(\lambda,\mu) - \lambda W_\lambda(\lambda,\mu) \\
0 &= -\lambda W_\mu(\lambda,\mu) - C_\mu(\mu,\theta) - \xi\alpha(\theta,\theta_{i-1})C_{\theta\mu}(\mu,\theta).
\end{aligned}
$$

Let $\lambda^*(\theta, \theta_{i-1})$ and $\mu^*(\theta, \theta_{i-1})$ be the solutions to the first-order conditions; then, with a fixed $MON(s)$ the expected organizational net value equals:

$$
\begin{aligned}
E\{NV(\theta); s\} &= \sum_{i=1}^{s} \int_{\theta_{i-1}}^{\theta_i} \{V(\lambda^*(\theta,\theta_{i-1})) - \lambda^*(\theta,\theta_{i-1})W(\lambda^*(\theta,\theta_{i-1}),\mu^*(\theta,\theta_{i-1})) \\
&\quad -C(\mu^*(\theta,\theta_{i-1}),\theta) - \xi\alpha(\theta,\theta_{i-1})C_\theta(\mu^*(\theta,\theta_{i-1}),\theta)\}\, dF(\theta). \quad (6.15)
\end{aligned}
$$

Example 6.2: Equally Spaced Partition Monitoring Systems. Assume that the assumptions made in Example 6.1 hold and that the monitoring system $MON(s)$ partitions Θ evenly so that $\theta_i - \theta_{i-1} = \frac{\bar{\theta}-\underline{\theta}}{s}$ for all $i \in \{1, 2, \ldots, s\}$. As shown previously, for all $\theta \in \Theta_i$,

$$
\begin{aligned}
\lambda^*(\theta,\theta_{i-1}) &= \left[\frac{k - \sqrt{\gamma^e(\theta,\theta_{i-1})}}{\gamma^e(\theta,\theta_{i-1})}\right]^2 \\
\mu^*(\theta,\theta_{i-1}) &= \frac{k\left[k - \sqrt{\gamma^e(\theta,\theta_{i-1})}\right]}{\gamma^e(\theta,\theta_{i-1})^2} \\
H^e(\theta,\theta_{i-1}) &= \gamma^e(\theta,\theta_{i-1})\lambda^*(\theta,\theta_{i-1}),
\end{aligned}
$$

where $\gamma^e(\theta,\theta_{i-1}) = \theta + \xi(\theta - \theta_{i-1}) = (1+\xi)\theta - \xi(i-1)r(s) - \xi\underline{\theta}$. The expected organizational net value then is:

$$
E\{NV^e(\theta); s\} = \sum_{i=1}^{s} \int_{\Theta_i} H^e(\theta,\theta_{i-1})dF(\theta)
$$

$$= \sum_{i=1}^{s} \int_{\underline{\theta}+(i-1)r(s)}^{\underline{\theta}+ir(s)} \frac{\left[k - \sqrt{\gamma^e(\theta, \underline{\theta} + (i-1)r(s))}\right]^2}{\gamma^e(\theta, \underline{\theta} + (i-1)r(s))} dF(\theta).$$

Figure 6.2 depicts the expected organizational net value for the equal-spaced $MON(s)$ with $k = 5$, $\xi = 0.5$, and $\Theta = [1, 2]$ as s ranges from 1 to 10.

The Optimal Partition of $MON(s)$ Systems

Given the number of intervals over Θ that a monitoring system is capable of partitioning, the optimal way for the central management to construct the partition needs to be identified. Hence the central management needs to solve:

$$E\{NV^*(\theta); s\} \equiv \max_{\theta_i} \sum_{i=1}^{s} \int_{\theta_{i-1}}^{\theta_i} H(\theta, \theta_{i-1}) dF(\theta) \tag{6.16}$$

subject to

$$\theta_i \geq \theta_{i-1}, \qquad \forall i \in \{1, \ldots, s\}. \tag{6.17}$$

Given a $MON(s)$ system, the optimal partition is never degenerate as long as $\xi > 0$. That is, letting $\theta_i^*, i \in \{1, \ldots, s-1\}$ denote the partition of a $MON(s)$ system, then $\underline{\theta} < \theta_1^* < \cdots < \theta_{s-1}^* < \overline{\theta}$.

PROPOSITION 6.3 *Given a $MON(s)$ system and that $\xi > 0$, θ_i^* is strictly increasing in i, i.e., $\underline{\theta} < \theta_1^* < \cdots < \theta_{s-1}^* < \overline{\theta}$. Furthermore, there exists a unique partition if for all $\theta_i \in (\theta_{i-1}, \theta_{i+1})$,*

$$\frac{\partial H(\theta_i, \theta_{i-1})}{\partial \theta_i} < \frac{dNV(\theta_i)}{d\theta_i},$$

where

$$NV(\theta_i) \equiv H(\theta_i, \theta_i) - \int_{\theta_i}^{\theta_{i+1}} \xi C_\theta(\mu^*(\theta, \theta_i), \theta) d\theta,$$

the ex post organizational net value when $\theta = \theta_i$.

PROOF. Differentiating (6.15) with respect to θ_i for each $i \in \{1, 2, \ldots, s-1\}$ gives the following set of first-order conditions:

$$\begin{aligned}
\frac{d}{d\theta_i} E\{NV(\theta); s\} &= [H(\theta_i, \theta_{i-1}) - H(\theta_i, \theta_i)] f(\theta_i) + \int_{\theta_i}^{\theta_{i+1}} \frac{dH(\theta, \theta_i)}{d\theta_i} dF(\theta) \\
&= [H(\theta_i, \theta_{i-1}) - H(\theta_i, \theta_i)] f(\theta_i) + \int_{\theta_i}^{\theta_{i+1}} \frac{\xi f(\theta_i) C_\theta(\mu^*(\theta, \theta_i), \theta)}{f(\theta)} dF(\theta) \\
&= 0,
\end{aligned}$$

where

$$\begin{aligned}
H(\theta_i,\theta_{i-1}) &= V(\lambda^*(\theta_i,\theta_{i-1})) - \lambda^*(\theta_i,\theta_{i-1})W(\lambda^*(\theta_i,\theta_{i-1}),\mu^*(\theta_i,\theta_{i-1})) \\
&\quad -C(\mu^*(\theta_i,\theta_{i-1}),\theta_i) - \xi\alpha(\theta_i,\theta_{i-1})C_\theta(\mu^*(\theta_i,\theta_{i-1}),\theta_i) \\
H(\theta_i,\theta_i) &= V(\lambda^*(\theta_i,\theta_i)) - \lambda^*(\theta_i,\theta_i)W(\lambda^*(\theta_i,\theta_i),\mu^*(\theta_i,\theta_i)) \\
&\quad -C(\mu^*(\theta_i,\theta_i),\theta_i) - \xi\alpha(\theta_i,\theta_i)C_\theta(\mu^*(\theta_i,\theta_i),\theta_i) \\
&= V(\lambda^*(\theta_i,\theta_i)) - \lambda^*(\theta_i,\theta_i)W(\lambda^*(\theta_i,\theta_i),\mu^*(\theta_i,\theta_i)) \\
&\quad -C(\mu^*(\theta_i,\theta_i),\theta_i).
\end{aligned}$$

After cancelling out $f(\theta)$ and noting that $f(\theta_i) > 0$, for $m = 1,\ldots,s-1$, the first-order conditions reduce to:

$$H(\theta_i,\theta_{i-1}) - H(\theta_i,\theta_i) + \int_{\theta_i}^{\theta_{i+1}} \xi C_\theta(\mu^*(\theta,\theta_i),\theta)d\theta = 0. \tag{6.18}$$

Fix θ_{i-1} and θ_{i+1} such that $\theta_{i-1} < \theta_{i+1}$. Since

$$\lim_{\theta_i \uparrow \theta_{i+1}} \int_{\theta_i}^{\theta_{i+1}} \xi C_\theta(\mu^*(\theta,\theta_i),\theta)d\theta = 0$$

and

$$H(\theta_i,\theta_{i-1})|_{\theta_i=\theta_{i+1}} < H(\theta_i,\theta_i)|_{\theta_i=\theta_{i+1}},$$

then

$$\frac{d}{d\theta_i}E\{NV(\theta);s\} < 0$$

when θ_i is in a neighborhood of θ_{i+1}. Similarly, if $\theta_i = \theta_{i-1}$,

$$\int_{\theta_i}^{\theta_{i+1}} \xi C_\theta(\mu^*(\theta,\theta_i),\theta)d\theta > 0.$$

Because

$$\lim_{\theta_i \downarrow \theta_{i-1}} H(\theta_i,\theta_{i-1}) = H(\theta_i,\theta_i),$$

$$\frac{d}{d\theta_i}E\{NV(\theta);s\} > 0$$

when θ_i is in a neighborhood of θ_{i-1}, and therefore there must be a $\theta_i^* \in (\theta_{i-1},\theta_{i+1})$ where

$$\int_{\Theta_i} H(\theta,\theta_{i-1})\, dF(\theta) + \int_{\Theta_{i+1}} H(\theta_{i+1},\theta)\, dF(\theta)$$

reaches its maximum. Since this is true for any two arbitrary intervals, θ_i^* must be strictly increasing in m.

The second part of the proposition is obvious. Since $f(\theta) > 0$ and

$$NV(\theta_i, \theta_i) = H(\theta_i, \theta_i) - \int_{\theta_i}^{\theta_{i+1}} \xi C_\theta(\mu^*(\theta, \theta_i), \theta) d\theta,$$

if $\frac{\partial H(\theta_i, \theta_{i-1})}{\partial \theta_i} < \frac{dNV(\theta_i, \theta_i)}{d\theta_i}$, $E\{H; s\}$ has exactly one critical point within $(\theta_{i-1}, \theta_{i+1})$ (i.e., $\frac{d}{d\theta_i} E\{NV(\theta); s\}$ changes its sign exactly once over $(\theta_{i-1}, \theta_{i+1})$), and therefore the first-order conditions (6.18) give the unique global maximum. ||

Interpreting the first-order condition (6.18) is straightforward. Given an interval $[\theta_{i-1}, \theta_{i+1}]$, an increase in θ_i moves some states into Θ_i. As a result, at the margin, the virtual organizational net value is increased by an amount equal to the sum of $H(\theta_i, \theta_{i-1})$ and the informational rent that state θ_i will generate for the IS manager when it serves as a lower bound in Θ_{i+1},

$$\xi \int_{\theta_i}^{\theta_{i+1}} C_\theta(\mu^*(\theta, \theta_i), \theta) d\theta.$$

But in so doing, the organization also loses some virtual organizational net value that is originally evaluated as the most efficient state in Θ_{i+1}. At optimality, these two effects must be balanced.

From the first part of Proposition 6.3, $\frac{d}{d\theta_i} E\{NV(\theta); s\}$ changes its sign at least once from positive to negative. So $\frac{d}{d\theta_i} E\{NV(\theta); s\}$ changes its sign exactly once, (6.18) gives the unique optimal partition. But for many cases both the virtual organizational net value in Θ_i and the true organizational net value in Θ_{i+1} are decreasing in θ_i. For instance, if $\alpha(\theta_i, \theta_{i-1})$ is non-decreasing in θ_i and $C_{\theta\theta}(\mu, \theta) \geq 0$, both $H(\theta_i, \theta_{i-1})$ and $NV(\theta_i)$ are decreasing in θ_i, since $\frac{\partial \mu^*(\theta, \theta_i)}{\partial \theta_i} > 0$ and

$$\frac{dNV(\theta_i)}{d\theta_i} = -(1-\xi) C_\theta(\mu^*(\theta_i, \theta_i), \theta_i) - \int_{\theta_i}^{\theta_{i+1}} C_{\theta\mu}(\mu^*(\theta, \theta_i), \theta) \frac{\partial \mu^*(\theta, \theta_i)}{\partial \theta_i} d\theta.$$

Thus, when both $H(\theta_i, \theta_{i-1})$ and $NV(\theta_i, \theta_i)$ are decreasing, the set of first-order conditions (6.18) gives the optimal partition if $H(\theta_i, \theta_{i-1})$ is decreasing faster than $NV(\theta_i, \theta_i)$ is in θ_i.

Example 6.3: The Optimal Partition of *MON*(s) Systems. When the assumptions in Example 6.1 hold, for each interval Θ_i,

$$\begin{aligned}\lambda^*(\theta,\theta_{i-1}) &= \left[\frac{k-\sqrt{\gamma(\theta,\theta_{i-1})}}{\gamma(\theta,\theta_{i-1})}\right]^2\\ \mu^*(\theta,\theta_{i-1}) &= \frac{k\left[k-\sqrt{\gamma(\theta,\theta_{i-1})}\right]}{\gamma(\theta,\theta_{i-1})^2}\\ H(\theta,\theta_{i-1}) &= \gamma(\theta,\theta_{i-1})\lambda^*(\theta,\theta_{i-1}),\end{aligned}$$

where $\gamma(\theta,\theta_{i-1}) \equiv \theta + \xi(\theta-\theta_{i-1})$, and the expected organizational net value is

$$E\{NV(\theta);s\} = \sum_{i=1}^{s}\int_{\Theta_i} H(\theta,\theta_{i-1})\,dF(\theta).$$

Then, given s, the first-order conditions are: for $i = 1,2,\ldots,s-1$,

$$\begin{aligned}\frac{d}{d\theta_i}E\{NV(\theta);s\} &= H(\theta_i,\theta_{i-1}) - H(\theta_i,\theta_i) + \xi\int_{\theta_i}^{\theta_{i+1}}\mu^*(\theta,\theta_i)d\theta\\ &= H(\theta_i,\theta_{i-1}) - \frac{H(\theta_i,\theta_i)+\xi H(\theta_{i+1},\theta_i)}{1+\xi}\\ &= 0.\end{aligned}$$

To verify that this set of first-order conditions yields the unique optimal partition, the second-order conditions must be checked. It is straightforward to show that

$$\frac{d^2}{d\theta_i^2}E\{NV(\theta);s\} = -(1+\xi)\,\mu^*(\theta_i,\theta_{i-1}) + \frac{\mu^*(\theta_i,\theta_i)-\xi^2\mu^*(\theta_{i+1},\theta_i)}{1+\xi}.$$

So it is sufficient to show that

$$(1+\xi)\,\mu^*(\theta_i,\theta_{i-1}) > \frac{1}{1+\xi}\,\mu^*(\theta_i,\theta_i)$$

for all θ_i and $\xi > 0$. But the above inequality is equivalent to

$$\frac{k-\sqrt{\theta_i+\xi\,(\theta_i-\theta_{i-1})}}{[\theta_i+\xi\,(\theta_i-\theta_{i-1})]^2} > \frac{k-\sqrt{\theta_i}}{[(1+\xi)\,\theta_i]^2},$$

or

$$\frac{k-\sqrt{\theta_i}}{k-\sqrt{\theta_i+\xi\,(\theta_i-\theta_{i-1})}} < \frac{[(1+\xi)\,\theta_i]^2}{[\theta_i+\xi\,(\theta_i-\theta_{i-1})]^2}.$$

It is then obvious that, as long as $\xi > 0$ and $\theta_i \geq \theta_{i-1} > 0$, the left-hand side of the inequality is less than or equal to one and the right-hand side of the inequality is strictly greater than one. Thus the first-order conditions are necessary and sufficient to solve for the unique optimal partition. Figure 6.3 shows an example of the optimal partitions for a $MON(s)$ system as s varies from 2 to 10. Figure 6.4 depicts the expected organizational net values that can be generated by equally spaced and optimally partitioned monitoring systems for cases with $\xi = 0.1$, 0.5, and 1. Clearly, for this example, an equally spaced monitoring system can perform almost as well as the optimal monitoring system. This is particularly the case when the IS manager's incentive problem is less severe (i.e., when ξ is small.) For instance, when $\xi = 0.1$, the expected organizational net values under the two systems nearly coincide. It is also clear that the effect of the IS manager's incentive problem declines as the central management's monitoring technology improves (i.e., as s increases.) With no monitoring, the difference between each case is larger than 2. But with $s = 10$, the difference is smaller than 0.5 even between the case with $\xi = 0.1$ and the case with $\xi = 1$.

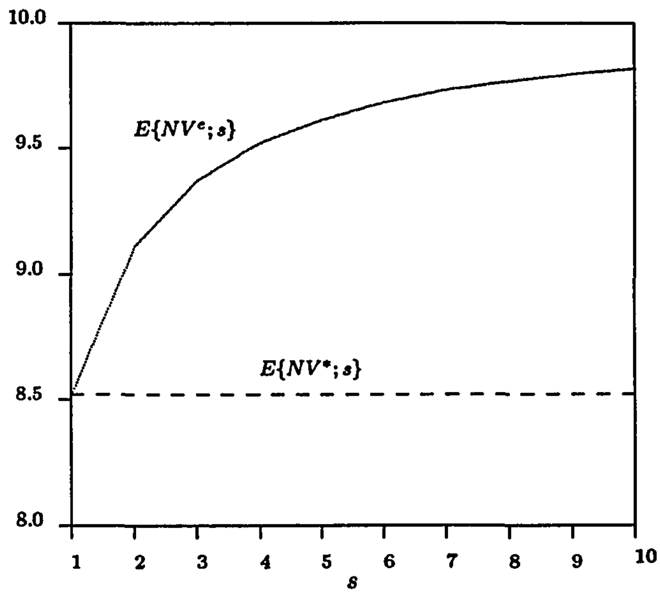


FIGURE 6.2: EXAMPLE 6.2—AN EXAMPLE OF AN EQUALLY SPACED SPACE-PARTITION MONITORING SYSTEM.

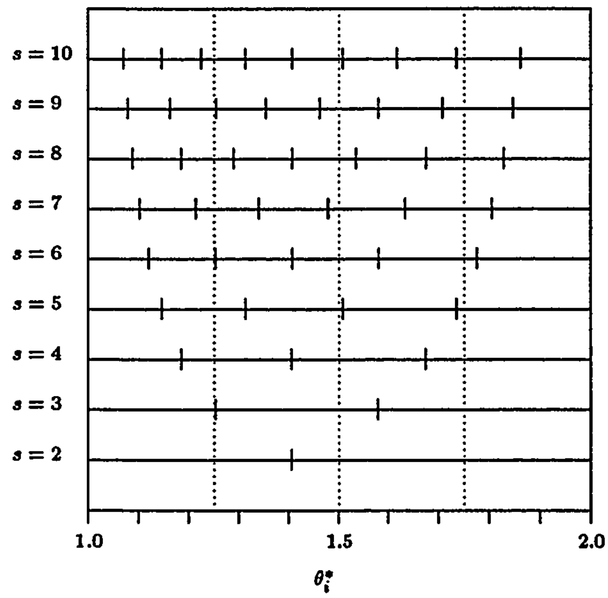
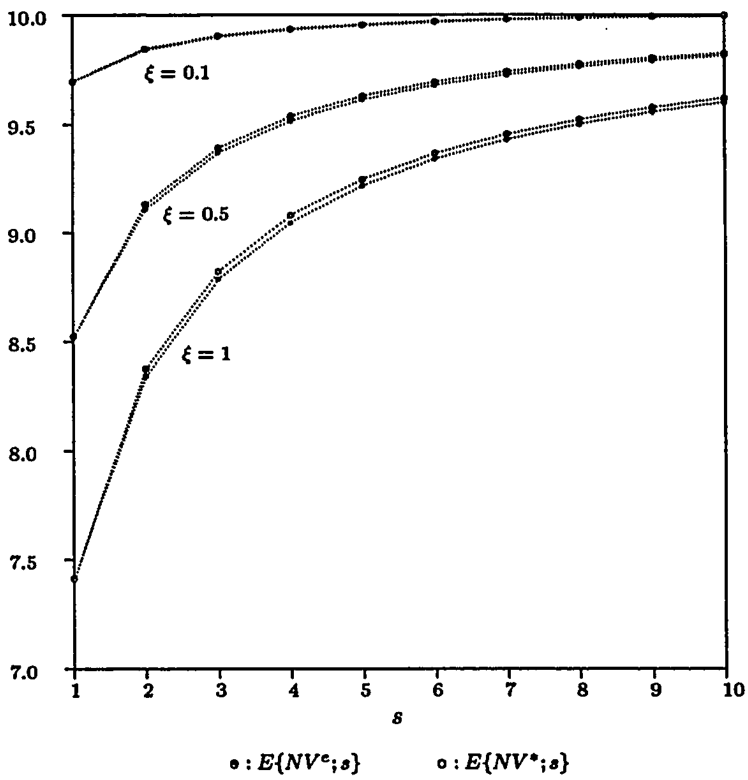


FIGURE 6.3: EXAMPLE 6.2—THE OPTIMAL PARTITION OF SPACE-PARTITION MONITORING SYSTEMS AS s VARIES FROM 2 TO 10.



The expected organizational net values with an equally spaced monitoring system ($E\{NV^e; s\}$) and the optimal monitoring system ($E\{NV^*; s\}$) as s varies from 1 to 10.

FIGURE 6.4: EXAMPLE 6.2—THE SPACE-PARTITION MONITORING SYSTEMS.

Chapter 7

Concluding Remarks and Future Research Directions

Perhaps the most fundamental contribution of this dissertation is the modeling, which combines a queuing model of the information system, a necessary ingredient for capturing its cost behavior, with a mechanism design approach to handle the cost information asymmetry problem between IS management and central management. This is an important foundation for raising the level of rigor with which IS management problems can be studied in the material on extensions and applications of the dissertation.

There are several implications of the results derived in the dissertation.

1. It was shown that the presence of those imperfections like information asymmetry and objective conflicts leads to reduced capacity, number of jobs served and utilization rate, higher prices, and longer mean waiting times. Thus organizations suffer losses not only from the IS manager's informational rent, but also from the opportunity cost of jobs not served. Consequently, recognizing only the incentive and information problems of the demand side is insufficient for deriving an effective mechanism to control IS resources; central management must also recognize the potential incentive and information problems of the supply side and try to alleviate them.

2. With unlimited communication, the revelation principle guarantees that, in the

presence of cost information asymmetry and incentive conflicts, a cost center governed by the optimal mechanism will be superior to any other organizational alternative for IS services by the criterion of the expected net value of the resource. The examples and numerical results provided in Chapters 3 and 4 give evidence that the difference in performance between the optimal mechanism and two practical competitors, a profit center and a naively governed cost center, is indeed significant. This coincides with the observed (low) frequency of profit centers in actual organizations (McGee [69]).

3. Central management must recognize that organizational slack is the inevitable result of information asymmetry, and its complete elimination is in fact undesirable. This was clearly seen in the case of the profit center: it is in fact very effective in limiting the IS manager's informational rent, but it also yielded much the worst organizational net value. The results for the optimal mechanism show that all the parties (the users, the IS manager, and the organization) will nearly always be better off under the mechanism. However, my results suggest that organizational performance should on average be better with direct compensation for reported cost savings rather than allowing the IS manager to consume all of the slack. Thus organizations with such incentive schemes should have better performing IS departments than those without them.

4. It is apparent from the mechanism design that the central management has an important role in the successful operation of an IS department. The most prominent ingredients are: (1) information on the possible range of capacity costs; (2) an idea of the degree of incentive conflict between the IS manager and central management; (3) the ability of the central management to make credible commitments; and (4) information about the gross value of computing to the organization. It is worth noting that elevating the head of the IS group to a central management position such as CIO will be likely to have a beneficial effect on all of these. In particular, regarding point (2), such a step can have the dual effect of helping to align incentives, thereby reducing the degree of conflict in the first place and also

of providing others in the central management with better information about the preferences of the IS manager. Furthermore, since the magnitude of the loss is increasing in the scale of the IS resources, as can be seen by the role of capacity in the expressions for the informational rent, one would expect such arrangements to be most common in larger organizations.

5. When an organization's communication channels and its central office's knowledge about IS operations are limited, the central management may want to consider decentralizing the IS-related decisions by organizing the IS department as a profit center. Although this will encourage the IS department to behave like a monopolist, the gains in flexibility from decentralization may outweigh the loss of control after taking into account the central management's limited ability to prescribe an effective control scheme and the required communication costs for supporting it.

6. Given the information problem, one would expect to see firms that offer independent assessments of IS department costs, and indeed there are such firms (Carlyle [16]). However, as Carlyle reports, the firms providing such information services have considerable difficulty attracting new clients, and their fees seem surprisingly low ($20,000 to $150,000 depending on the size of the department). One possible explanation for this can be seen from the comparative statics of mean-preserving spreads in Chapter 3, which correspond to information that reduces the central management's uncertainty about the range of costs without changing the expected value of costs. A reduction in the spread could either increase or decrease the expected net value of a cost center, and central management may therefore be unwilling to pay for such information. Since in general the information may both reduce uncertainty and alter the expected value, it is clear that the information's expected value is quite difficult to assess and likely to be quite low, causing few firms to use the services, and those that do use them to be willing to pay relatively little. Moreover, as demonstrated in Chapter 6, depending on the monitoring technology, a truth-revelation mechanism may not be optimal. This also manifests the complexity of determining an appropriate control mechanism

in a more realistic setting as the central management's ability to verify the IS department's report is enhanced.

The ultimate value of the dissertation will, I believe, be determined by the degree to which it enables other problems of IS management to be formally studied. Methodologically, mechanism design allows a number of other IS management problems to be studied more rigorously than was previously possible. For example, it is very useful in analyzing bargaining situations, such as contracting for software development, and in analyzing relationship-specific IS investments. The revelation principle not only generates the welfare upper-bound for the mechanism designer but can make complex problems more tractable as well. Mechanism design can also be extended to multi-agent settings, so that it is possible to study IS resource allocation problems, such as controlling networks and distributed computing as well as issues in downsizing and system integration.

More specifically, I consider in this dissertation only the case in which neither the users nor the IS department has access to the external market, so one direct extension of my present model would consider cases with potential external service providers to which the organization could outsource its information processing activities. Using a market as an incentive scheme has been studied in different principal-agent models, e.g., Hart [47] and Scharfstein [95]. The role of market forces and competition in regulating public firms and government procurement has also been studied extensively, e.g., Anton and Yao [4, 5], Caillaud [13], Demski et al. [24], Lewis and Sappington [64], and Riordan and Sappington [93].

To illustrate the potential impact of external competition, consider the case studied in Chapter 3. Assume that there exists an external service provider who offers services similar to those provided by the internal IS department. Further assume that the central management has decided to outsource either all or none of the organization's IS operations. Without explicitly considering the required negotiation process, let NV^e be the common knowledge expected organizational net value that the organization can obtain from outsourcing. As in Chapter 3, let $H(\theta)$ be the virtual organizational net value, which is strictly decreasing in θ. By pointwise comparison of the benefits and costs, it is clear that if there exist a $\theta^* \in \Theta$ such that $H(\theta^*) = NV^e$, the organization will be better off to outsource whenever the IS department's reported parameter is higher than θ^*. When this is the case, NV^e in effect puts a lower bound on the organization's net value. The effect of an alternative source on limiting the IS department's informational rents can now be easily seen. Since θ^* puts an upper bound on the information that the IS manager can exaggerate, the excess budget allocation required for inducing truth-revelation is $\int_{\theta}^{\theta^*} C(\mu^*(\tilde{\theta}), \tilde{\theta})\, dF(\tilde{\theta})$ for all $\theta \leq \theta^*$. Consequently, the costs required to induce the IS manager's truth-revelation are reduced for all $\theta \leq \theta^*$. Therefore, whenever the internal IS department is retained, the scale of IS operations, μ, will be larger than when alternative sources are absent.

This simple example illustrates two obvious effects of external market forces. First, when the external service provider is very efficient, the organization's net value can be increased by outsourcing. Second, and more important, the central management can use the threat of outsourcing as an endogenous control mechanism to reduce the costs of ex ante incentives. As a result, even when the external service providers are not as efficient as the internal IS department, the costs of achieving a particular scale of IS operations without actually outsourcing can still be reduced.

Of course, in a more realistic setting, there are at least three other issues that need to be considered. First, when an organization has decided to explore the possibility of outsourcing by negotiating with external service providers or inviting them to bid for service contracts, the prices quoted by these service providers may give informative signals about the internal IS department's costs. The effects of such signals on limiting informational rents of a regulated public firm are studied by Caillaud [13] and Demski et al. [24]. Lazear and Rosen [62], Nalebuff and Stiglitz [87], and Shleifer [98] study the role of correlation between the environmental parameters of several regulated agents in a moral hazard framework. The issue then is to determine how the cost projections generated by the extra signals can be translated into production and compensation levels. Second, the internal IS department is usually much more informed about the organization's idiosyncratic system characteristics and operating environment. This kind of knowledge was developed possibly through years of relationship-specific investment and is difficult to transfer; its value may be demonstrated by the fact that external service providers willingly retain most of an organization's internal IS staffers even after they take over the organization's entire IS operations. Third, technological innovations may have an impact on an organization's outsourcing decision. For example, advanced telecommunication technologies and the industry trend toward open systems may increase the number of alternative sources that an organization can explore. Finally as indicated by Carlyle [15], IS spinouts tend to be evaluated as "profit-or-loss" centers by the organization. A natural extension of my current work would investigate the performance of IS organizations when access to the external market is permitted for both the users and the IS department.

Bibliography

[1] ALLEN, B. (1987) "Make Information Services Pay its Way," *Harvard Business Review*, 57-63.

[2] ANTLE, R. AND G. EPPEN (1985) "Organizational Slack in Capital Budgeting," *Management Science*, **31**, 163–174.

[3] ANTLE, R. AND J. FELLINGHAM (1990) "Resource Rationing and Organizational Slack in a Two-Period Model," *Journal of Accounting Research*, **28**, 1–24.

[4] ANTON, J. J. AND D. A. YAO (1987) "Second Sourcing and the Experience Curve: Price Competition in Defense Procurement," *Rand Journal of Economics*, **18**, 57–76.

[5] ANTON, J. J. AND D. A. YAO (1989) "Split Awards, Procurement, and Innovation," *Rand Journal of Economics*, **20**, 538–552.

[6] BALINSKI, M. L. AND F. M. SAND (1985) "Auctioning Landing Rights at Congested Airports," in H. P. Young, ed., *Cost Allocation: Methods, Principles, Applications*, Amsterdam: Elsevier Science.

[7] BARON, D. P. (1989) "Design of Regulatory Mechanisms and Institutions," in R. Schmalensee and R. D. Willig, ed., *Handbook of Industrial Organization*, Amsterdam: Elsevier Science.

[8] BARON, D. P. AND D. BESANKO (1984) "Regulation, Asymmetric Information, and Auditing," *Rand Journal of Economics*, **15**, 447–470.

[9] BARON, D. P. AND R. B. MYERSON (1982) "Regulating a Monopolist with Unknown Cost," *Econometrica*, **50**, 911–930.

[10] BARRON, T. (1992) "Some New Results in Testing for Economies of Scale in Computing: 1985 and 1988 Data," *Decision Support Systems, Special Issue on Information Systems and Economics*, **8**, 405–429.

[11] BELL, C. AND S. STIDHAM (1983) "Individual versus Social Optimization in the Allocation of Customers to Alternative Servers," *Management Science*, **29**, 831–839.

[12] BIDDLE, G. C. AND R. STEINBERG (1984) "Allocations of Joint and Common Costs," *Journal of Accounting Literature*, **3**, 1–45.

[13] CAILLAUD, B. (1990) "Regulation, Competition, and Asymmetric Information," *Journal of Economic Theory*, **52**, 87–110.

[14] CARLIN, A. AND R. E. PARK (1970) "Marginal Cost Pricing of Airport Runway Capacity," *American Economic Review*, **60**, 310-319.

[15] CARLYLE, R. E. (1989) "The Selling of IS," *Datamation*, July, 22–26.

[16] CARLYLE, R. E. (1990) "Getting a Grip on Costs," *Datamation*, July 15, 20–23.

[17] CASE L. S. AND L. B. LAVE (1977) "Tolls: Efficiency and Equity Issues for Inland Waterways," *Management Science*, **23**, 811–819.

[18] COUGER, J. D. (1990) "Motivating Analysts and Programmers," *Computerworld*, **24**, January 15, 75–76.

[19] CURIEN, N. (1985) "Cost Allocation and Pricing Policy: The Case of French Telecommunications," in H. P. Young, ed., *Cost Allocation: Methods, Principles, Applications*, Amsterdam: Elseviers Science.

[20] CYERT, R. AND J. MARCH (1963) *The Behavior Theory of the Firm*, Englewood Cliffs: Prentice-Hall.

[21] DASGUPTA, P. S., P. J. HAMMOND, AND E. S. MASKIN (1979) "The Implementation of Social Choice Rules: Some Results on Incentive Compatibility," *Review of Economic Studies*, **46**, 185–216.

[22] DEMSETZ, H. (1991) "The Theory of the Firm Revisited," in O. Williamson and S. Winter, ed., *The Nature of the Firm: Origins, Evolution, and Development*, Oxford: Oxford University Press.

[23] DEMSKI, J. S. AND D. E.M. SAPPINGTON (1986) "Line Item Reporting, Factor Acquisition, and Subcontracting," *Journal of Accounting Research*, **24**, 250–269.

[24] DEMSKI, J. S., D. E.M. SAPPINGTON, AND P. T. SPILLER (1987) "Managing Supplier Switching," *Rand Journal of Economics*, **18**, 77–97.

[25] DEWAN, S. AND H. MENDELSON (1990) "User Delay Costs and Internal Pricing for a Service Facility," *Management Science*, **36**, 1502–1517.

[26] DEWEES, D. (1979) "Estimating the Time Costs of Highway Congestion," *Econometrica*, **47**, 1944–1512.

[27] DIAMOND, P. A. (1973) "Consumption Externalities and Imperfect Corrective Pricing," *Bell Journal of Economics and Management Science*, **4**, 526–538.

[28] DOLAN, R. J. (1978) "Incentive Mechanisms for Priority Queueing Problems," *Bell Journal of Economics*, **9**, 421–436.

[29] DYE, R. A. (1986) "Optimal Monitoring Policies in Agencies," *Rand Journal of Economics*, **17**, 339–350.

[30] DYER, M. E. AND L. G. PROLL (1977) "On the Validity of Marginal Analysis for Allocation of Servers in $M/M/c$ Queues," *Management Science*, **23**, 1019–1022.

[31] EDELSON, N. M. AND D. K. HILDEBRAND, (1975) "Congestion Tolls for Poisson Queueing Process," *Econometrica*, **43**, 81–92.

[32] FUDENBERG, D. AND J. TIROLE (1991) *Game Theory*, Cambridge: MIT Press, MA.

[33] GENESERETH, M. R., M. L. GINSBERG, AND J. S. ROSENCHEIN (1986) "Cooperation without Communication," *Proc. AAAI-86*, 561-567.

[34] GEORGEFF, M. P. (1983) "Communication and Interaction in Multi-agent Planning," *Proc. AAAI-83*, 125–129.

[35] GEORGEFF, M. P. (1986) "The Representation of Events in Multiagent Domains," *Proc. AAAI-86*, 70–75.

[36] GRASSMANN, W. (1983) "The Convexity of the Mean Queue Size of the $M/M/c$ Queue with Respect to the Traffic Intensity," *Journal of Applied Probability*, **20**, 916–919.

[37] GREEN, J. R. AND J.-J. LAFFONT (1986) "Partially Verifiable Information and Mechanism Design," *Review of Economic Studies*, **53**, 447–456.

[38] GREEN, J. R. AND J.-J. LAFFONT (1986) "Incentive Theory with Data Compression," in W. P. Heller et al. ed., *Uncertainty, Information, and Communication: Essays in Honor of Kenneth J. Arrow, Vol III*, Cambridge: Cambridge University Press.

[39] GREEN, J. R. AND J.-J. LAFFONT (1986) "Alternative Limited Communication Systems: Centralization versus Interchange of Information," in W. P. Heller et al. ed., *Uncertainty, Information, and Communication: Essays in Honor of Kenneth J. Arrow, Vol III*, Cambridge: Cambridge University Press.

[40] GREEN, J. R. AND J.-J. LAFFONT (1987) "Limited Communication and Incentive Compatibility," in T. Groves et al., ed., *Information, Incentive and Economic Mechanisms*, University of Minnesota Press.

[41] GREEN, J. R. AND J.-J. LAFFONT (1987) "Posterior Implementability in a Two-person Decision Problem," *Econometrica*, **55**, 69–94.

[42] GROVES, T. (1973) "Incentive in Teams," *Econometrica*, **41**, 617–663.

[43] GROVES, T. (1985) "The Impossibility of Incentive Compatible and Efficient Full Cost Allocation Scheme," in H. P. Young, ed., *Cost Allocation: Methods, Principles, Applications*, Amsterdam: Elseviers Science.

[44] GROVES, T. AND M. LOEB (1975) "Incentives and Public Inputs," *Journal of Public Economics*, 4, 211-226.

[45] GUESNERIE, R. AND J.-J. LAFFONT (1984) "A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-Managed Firm," *Journal of Public Economics*, **25**, 329–369.

[46] HARRIS, M. AND R. M. TOWNSEND (1981) "Resource Allocation under Asymmetric Information," *Econometrica*, **49**, 33–64.

[47] HART, O. (1983) "The Market Mechanism as an Incentive Scheme," *Bell Journal of Economics*, 14, 366-382.

[48] HART, O. (1991) "Incomplete Contract and the Theory of the Firm," in O. Williamson and S. Winter, ed., *The Nature of the Firm: Origins, Evolution, and Development*, Oxford: Oxford University Press.

[49] HEYMAN, D.P. AND M.J. SOBEL (1982) *Stochastic Models in Operations Research Vol I*, McGraw-Hill.

[50] HIRSHLEIFER, J. (1956) "On the Economics of Transfer Pricing," *Journal of Business*, July, 172-184.

[51] HIRSHLEIFER, J. (1958) "Peak Loads and Efficient Pricing: Comment," *Quarterly Journal of Economics*, **58**, 451–462.

[52] HOLMSTROM, B. (1979) "Moral Hazard and Observability," *Bell Journal of Economics*, **10**, 74–91.

[53] HOLMSTROM, B. AND R. B. MYERSON (1982) "Efficient and Durable Decision Rules with Incomplete Information," *Econometrica*, **51**, 1799–1819.

[54] KAPLAN, R. S. AND A. A. ATKINSON (1989) *Advanced Management Accounting*, 2nd ed. Englewood Cliffs: Prentice-Hall.

[55] KEELER, T. E. AND K. SMALL (1977) "Optimal Peak-load Pricing, Investment, and Service Levels on Urban Expressways," *Journal of Political Economy*, **85**, 1-25.

[56] KEEN, P. (1991) "End of IS Budgets as We Know Them?" *Computerworld* special double issue, 24, 52–53, January 1, 15.

[57] KIRBY, A. J. ET AL. (1991) "Participation, Slack, and Budget-Based Performance Evaluation," *Journal of Accounting Research*, **29**, 109–128.

[58] KIRKPATRICK, D. (1991) "Why Not Farm Out Your Computing," *Fortune*, September 23, 103–112.

[59] KNUDSEN, N. C. (1972) "Individual and Social Optimization in a Multi-server Queue with a General Cost-Benefit Structure," *Econometrica*, **40**, 515–528.

[60] KOFMAN, F. AND J. RATLIFF (1991) "Johnny Carson vs. the Smothers Brothers: Monolog vs. Dialog in Costly Bilateral Communication," Working Paper #3254-91-EFA, Sloan School of Management, MIT.

[61] LAFFONT, J.-J. AND J. TIROLE (1986) "Using Cost Observation to Regulate Firms," *Journal of Political Economy*, **94**, 614–641.

[62] LAZEAR, E. AND S. ROSEN (1981) "Rank Order Tournament as Optimal Labor Contracts," *Journal of Political Economics*, **89**, 841–864.

[63] LEE, H. L. AND M. A. COHEN (1983) "A Note on the Convexity of Performance Measures of $M/M/c$ Queueing Systems," *Journal of Applied Probability*, **20**, 920-923.

[64] LEWIS, T. R. AND D. E.M. SAPPINGTON (1991) "On the Boundaries of the Firm: Self Provision vs. Subcontracting," *American Economic Review*, **81**, 887-900.

[65] LIPPMAN, S. A. (1975) "Applying a New Device in the Optimization of Exponential Queueing Systems," *Operations Research*, **23**, 687–710.

[66] LIPPMAN, S. A. AND S. STIDHAM (1977) "Individual Versus Social Optimization in Exponential Congestion Systems," *Operations Research*, **25**, 233–247.

[67] LITTLECHILD, S. C. AND G. F. THOMPSON, (1977) "Aircraft Landing Fees: A Game Theory Approach," *Bell Journal of Economics*, **8**, 186-204.

[68] MAGLITTA, J. (1991) "Squeezing More from IS Dollars," *Computerworld* special double issue, 24, 52–53, January 1, 2–4.

[69] MCGEE, R. W. (1988) *Accounting for Data Processing Costs*, Greenwood Press.

[70] MCAFEE, R. P. AND J. MCMILLAN (1987) "Competition for Agency Contracts," *Rand Journal of Economics*, **18**, 296–307.

[71] MELUMAD, N., D. MOOKHERJEE, AND S. REICHELSTEIN (1990) "Hierarchical Decentralization of Incentive Contracts," mimeo, Stanford University.

[72] MENDELSON, H. (1985) "Pricing Computer Services: Queueing Effects," *Communication of the ACM*, **28**, 312–321.

[73] MENDELSON, H. (1987) "Economies of Scale in Computing: Grosch's Law Revisited," *Communication of the ACM*, **30**, 1066–1072.

[74] MENDELSON, H. (forthcoming) *The Economics of Information Systems Management*, Englewood Cliffs: Prentice Hall.

[75] MENDELSON, H. AND U. YECHIALI (1981) "Controlling the $GI/M/1$ Queue by Conditional Acceptance Customers," *Eurapean Journal of Operations Research*, **7**, 77–85.

[76] MENDELSON, H. AND S. WHANG (1990) "Optimal Incentive-Compatible Priority Pricing for the $M/M/1$ Queue," *Operations Research*, **38**, 870–883.

[77] MILGROM, P. R. (1982) "Good News and Bad News: Representation Theorems and Application," *The Bell Journal of Economics*, **12**, 380–391.

[78] MIRRLEES, J. (1986) "The Theory of Optimal Taxation," in K. J. Arrow and M. D. Intrilligator, ed., *Handbook of Mathematical Economics, Vol III*, Amsterdam: Elseviers Science.

[79] MOHRING, H. (1970) "The Peak Load Problem with Increasing Returns and Pricing Constraints," *American Economic Review*, **60**, 693-705.

[80] MOHRING, H. (1976) *Transportation Economics*, Cambridge: Ballinger, MA.

[81] MOHRING, H. AND M. HARWITZ (1970) *Highway Benefits: An Analytical Framework*, Evanston: Northwestern University Press, 84–86.

[82] MORRISON, S. A. (1983) "Estimation of Long-run Prices and Investment Levels for Airport Runways," *Research in Transportation Economics*, **1**, 103–130.

[83] MYERSON, R. B. (1979) "Incentive Compatibility and the Bargaining Problem," *Econometrica*, **47**, 61–73.

[84] MYERSON, R. B. (1985) "Analysis of Two Bargaining Problems with Incomplete Information," in A. E. Roth, ed., *Game-theoretic Models of Bargaining*, Cambridge: Cambridge University Press, 115–147.

[85] MYERSON, R. B. (1985) "Bayesian Equilibrium and Incentive Compatibility: An Introduction," in L. Hurwicz et al., ed., *Social Goals and Social Organization*, Cambridge: Cambridge University Press, 229–259.

[86] MYERSON, ROGER B. (1991) *Game Theory: Analysis of Conflict*, Cambridge: Harvard University Press, MA.

[87] NALEBUFF, B. AND J. STIGLITZ (1983) "Prizes and Incentives: Toward a General Theory of Compensation and Competition," *Bell Journal of Economics*, **14**, 166-178.

[88] NAOR, P. (1969) "On the Regulation of Queue Size by Levying Tolls," *Econometrica*, **37**, 15-24.

[89] OLIVELLA, P. (1989) "Information Structures and the Delegation of Monitoring," Discussion Paper No. 839, Economics Department, Northwestern University.

[90] ONIKI, H. (1986) "The Cost of Communication in Economic Organization: II," in W. P. Heller et al., ed., *Uncertainty, Information, and Communication: Essays in Honor of Kenneth J. Arrow, Vol.III*, Cambridge: Cambridge University Press.

[91] ORON, Y., D. PINES, AND E. SHESHINESKI, (1973) "Optimum vs. Equilibrium Land Use Pattern and Congestion Toll," *Bell Journal of Economics and Management Science*, 4, 619–636.

[92] OUM, T. H. AND Y, ZHANG (1987) "Airport Pricing: Congestion Tolls, Lumpy Investment, aand Cost Recovery," *Journal of Public Economics*, **43**, 353–374.

[93] RIORDAN, M. H. AND D. E.M. SAPPINGTON, (1989) "Second Sourcing" *Rand Journal of Economics*, **20**, 41–58.

[94] ROSENSCHEIN, J. S. (1986) "Rational Interaction: Cooperation Among Intelligent Agents," Ph.D. Thesis, Stanford University.

[95] SCHARFSTEIN, D. (1988) "Product Market Competition and Managerial Slack," *Rand Journal of Economics*, **19**, 147–155.

[96] SHARKEY, W. W. (1982) "Suggestions for a Game-theoretic Approach to Public Utility Pricing and Cost Allocation," *Bell Journal of Economics*, **13**, 57-68.

[97] SHARKEY, W. W. (1985) "Economic and Game-Theoretic Issues Associated with Cost Allocation in a Telecommunication Network," in H. P. Young, ed., *Cost Allocation: Methods, Principles, Applications*, Amsterdam: Elseviers Science.

[98] SHLEIFER, A. (1985) "A Theory of Yardstick Competition," *Rand Journal of Economics*, **16**, *319–327.*

[99] SILBERBERG, E. (1990) *The Structure of Economics: A Mathematical Analysis*, 2nd ed., New York: McGraw-Hill.

[100] SIMON, H. A. (1961) *Administrative Behavior*, 2nd ed., New York: Macmillan.

[101] Singh, N. (1985) "Monitoring and Hierarchies: The Marginal Value of Information in a Principle-Agent Model," *Journal of Political Economy*, **93**, 599–609.

[102] Solow, R. M. (1973) "Congestion Cost and the Use of Land for Streets," *Bell Journal of Economics and Management Science*, 4, 602–618.

[103] Squire, L. (1973) "Some Aspects of Optimal Pricing for Telecommunication," *Bell Journal of Economics and Management Science*, 4, 515–538.

[104] Stidham, S. (1992) "Pricing and Capacity Decisions for a Service Facility: Stability and Multiple Local Optima." *Management Science*, **38**, 1121–1139.

[105] Strotz, R. (1964) "Urban Transportation Parables," in J. Margolis, ed., *The Public Economy of Urban Communities*, Baltimore: John Hopkins University Press, 127–169.

[106] Tirole, J. (1988) *The Theory of Industrial Organization*, Cambridge: MIT Press, MA.

[107] Vickery, W. (1969) "Congestion Theory and Transportation Investment," *American Economic Review*, **59**, 251–260.

[108] Walters, A. A. (1968) "The Theory and Measurement of Private and Social Cost of Highway Congestion," *Econometrica*, **29**, 676–699.

[109] Wang, T. and U. Sumita (1992) "Price and Capacity Decisions for Jackson Open Network," mimeo, University of Rochester.

[110] Whang, S. (1988) "Pricing Computer Services: Incentive, Information and Queueing Effects," Ph.D. Dissertation, University of Rochester.

[111] Whang, S. (1989) "Cost Allocation Revisited: An Optimality Result," *Management Science*, **35**, 1264–1273.

[112] Whang, S. (1990) "Alternative Mechanisms of allocating Computer Resources under Queueing Delays," *Information Systems Research*, **1**, 71–88.

[113] WIDDER, D. V. (1961) *Advanced Calculus*, Englewood Cliffs: Prentice-Hall.

[114] WILDER, C. (1991) "New Outsourcing Deals for IBM" *Computerworld*, **25**, January 7, 12.

[115] WILLIAMSON, O. E. (1966) "Peak-load Pricing and Optimal Capacity under Indivisibility Constraints," *American Economic Review*, **56**, 810–827.

[116] WILLIAMSON, O. E. (1985) *The Economic Institutions of Capitalism*, New York: The Free Press.

[117] YECHIALI, U. (1971) "On Optimal Balking Rules and Toll Charges in the $GI/M/1$ Queue Process," *Operations Research*, **19**, 348–370.

[118] YOUNG, H. P., ed. (1985) *Cost Allocation: Methods, Principles, Application*, Amsterdam: Elseviers Science.

[119] ZIMMERMAN, J. L. (1979) "The Costs and Benefits of Cost Allocations," *The Accounting Review*, **54**, 504–521.