AD-A045 111 HARVARD UNIV CAMBRIDGE MASS DIV OF ENGINEERING AND --ETC F/G 12/2
INFORMATION AND SIGNALING IN DECENTRALIZED DECISION PROBLEMS.(U)
SEP 77 M P KASTNER N00014-75-C-0648

UNCLASSIFIED TR-669 NL

1 OF 2
AD A045 111

(12)

(15) Office of Naval Research
Contract N00014-75-C-0648, NR-372-012
National Science Foundation Grant NSF-ENG-76-11824

(6) INFORMATION AND SIGNALING IN DECENTRALIZED DECISION PROBLEMS.

By
(10) Marcia P. Kastner

(11) Sep[redacted]77

(12) 137 P.

(14) TR-669

(9) Technical Report, No. 669

Division of Applied Sciences
Harvard University Cambridge, Massachusetts

use → 163750

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report No. 669	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) INFORMATION AND SIGNALING IN DECENTRALIZED DECISION PROBLEMS		5. TYPE OF REPORT & PERIOD COVERED Interim Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Marcia P. Kastner		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0648 NSF ENG 76-11824
9. PERFORMING ORGANIZATION NAME AND ADDRESS Division of Applied Sciences Harvard University Cambridge, Massachusetts 02138		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE September 1977 13. NUMBER OF PAGES 136
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted by the U.S. Government.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Decentralized Decision Problems
Information Structures
Nonzero-Sum Games
Team Theory
Information Theory
Market Signaling

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A decentralized many-person decision problem is one where each decision maker has different information. If one decision maker's information depends on what another decision maker has done, the information is called dynamic. In the past, problems involving dynamic information have been very difficult, if not impossible, to solve. Two specific examples which have been solved, one from economic theory and the other from classical information theory, will be investigated. It will be shown that they can be formulated as two-person decision problems with the type of dynamic information structure called

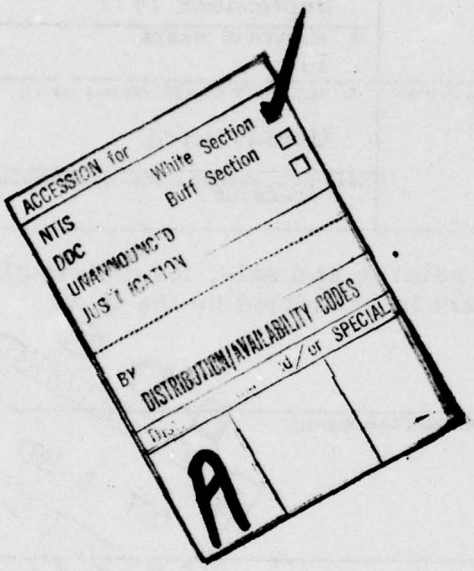
DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract continued

signaling. The first example involves a model of the job market as a non-zero-sum game. New equilibrium solutions are found and properties of these solutions, such as stability, multiple solutions, and threshold effects of signaling cost and noise, are studied. The second example models the Shannon problem as a team theory problem. The concept of real-time information theory is introduced, where source and channel sequences are of a fixed length, and general results about real-time solutions are proved and demonstrated.



Office of Naval Research

Contract N00014-75-C-0648 NR-372-012

National Science Foundation Grant ENG 76-11824

INFORMATION AND SIGNALING IN DECENTRALIZED DECISION PROBLEMS

By

Marcia P. Kastner

Technical Report No. 669

September 1977

The research reported in this document was made possible through support extended the Division of Applied Sciences, Harvard University by the U. S. Army Research Office, the U. S. Air Force Office of Scientific Research and the U. S. Office of Naval Research under the Joint Services Electronics Program by Contract N00014-75-C-0648 and by the National Science Foundation Grant ENG 76-11824 and in part during the tenure of the author as an IBM Predoctoral Fellow.

Division of Applied Sciences

Harvard University · Cambridge, Massachusetts

ABSTRACT

A decentralized many-person decision problem is one where each decision maker has different information. If one decision maker's information depends on what another decision maker has done, the information is called "dynamic." In the past, problems involving dynamic information have been very difficult, if not impossible, to solve. Two specific examples which have been solved, one from economic theory and the other from classical information theory, will be investigated. It will be shown that they can be formulated as two-person decision problems with the type of dynamic information structure called "signaling." The first example involves a model of the job market as a nonzero-sum game. New equilibrium solutions are found and properties of these solutions, such as stability, multiple solutions, and threshold effects of signaling cost and noise, are studied. The second example models the Shannon problem as a team theory problem. The concept of real-time information theory is introduced, where source and channel sequences are of a fixed length, and general results about real-time solutions are proved and demonstrated.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Information available to a decision maker (DM) not only influences his actions, but also determines whether a solution to the decision problem exists at all. The role of information becomes particularly complicated when there is more than one DM, especially if each DM has different information. In order to study how information influences decision making in many-person decision problems, two specific examples, one from economic theory and one from classical information theory, will be examined. It will be shown that they can be formulated as two-person decision problems. This formulation provides a framework for studying each problem's information structure, that is, "who knows what." The reason these particular examples have been chosen is because they both exhibit a special type of information structure called "signaling." In the past (see [1], [4], and [7]), problems involving this type of information structure have been very difficult, if not impossible, to solve. However, these two examples can be solved (in a sense to be defined). Thus, they provide new insights into possible new solution techniques.

Before going on to the two problems in detail, we first will define more precisely what is meant by a many-person decision problem with "signaling."

Suppose there are N decision makers, with the i-th DM denoted as DMi. First of all, let $x \in \Omega$ be a random variable representing the

state of the world (or state of nature) that each DM would like to know, with probability density $p(x)$. Secondly, for $i = 1, \ldots, N$, DMi's information $z_i \in Z_i$ is a function of x, written as $z_i = h_i(x)$.* When $h_i \neq h_j$ for $i \neq j$, then we say the problem is "decentralized," since each DM is making decisions based on different information. Thirdly, DMi's action, or decision, $u_i \in U_i$ is a function γ_i of his information, expressed as $u_i = \gamma_i(z_i)$, where γ_i is called a "strategy" or "decision rule." Lastly, DMi's objective, or payoff, function J_i is the expectation of a function $\mathcal{J}_i$ of x and all of the DMs' strategies. Each DMi now faces the following problem: choose a strategy γ_i from a class of specified admissible strategies (usually taken to be the class of measurable functions from Z_i to U_i) to minimize

$$J_i(\gamma_1, \ldots, \gamma_N) = E[\mathcal{J}_i(x, \gamma_1, \ldots, \gamma_N)] \quad .$$

The information to DMi can be modified to include not only z_i, a a measurement of the state x, but also the actions of the other decision makers. For example, suppose DMi's information also includes u_j, the action of DMj, $j \neq i$. Thus, a sense of order is conveyed in that DMj acts before DMi, and DMi observes this action. When this happens, that is, when DMi's information depends on what another person has done, we say that DMi has a <u>dynamic</u> information structure [2], [3]. Otherwise, the information structure is called <u>static</u>.**

---

*This is not the most general definition of information, but is sufficient for our purposes at this time.

**We are considering only <u>nonclassical</u> information in that each decision maker has different information [8].

The former is the type of information structure that occurs in the two examples to be studied.*

Sometimes a dynamic information structure can be reduced to a static one [2]. However, one example where this is not the case is the type of dynamic information structure we call "signaling." For the case of $N = 2$, let DM1's action be denoted as u and DM2's as v. Then signaling is defined as the type of dynamic information where DM1's information is just x and DM2's information is just u. In other words, DM1 "signals" his knowledge of x to DM2 through his action $u = \gamma_1(x)$. DM2 must now infer x from u and choose $v = \gamma_2(u)$.

Chapter II examines an economic application of signaling based on a model of the job market by Spence [6]. Although Spence used the term "signaling" to describe the type of information transfer in the job market, Chapter II extends the model by formulating the problem as a two-person decision problem. The reason for this is twofold: first of all, we immediately see that Spence's model is an example of a problem with (nonclassical) dynamic information that can be solved. For this reason, it provides an excellent vehicle for studying this type of information structure. Secondly, this set-up allows us to find new solutions and investigate different properties of the solutions. Although the lack of detail in the model prevents us from asserting the absolute validity of the economic issues raised, the decision- and control-theoretic framework provides qualitative insights into modeling the transfer of information.

---

*The reason problems with a dynamic information structure are difficult to solve is because the underlying probability distributions needed to find the solution are themselves solution-dependent. See [8] and [3] for details.

Chapter III deals with problems in Shannon theory, also sometimes referred to as classical information theory, which addresses the problem of coding a message and sending it through a noisy communication channel [5]. At first glance, this may sound unrelated to the economics-oriented Spence problem of Chapter II, but we will show that this problem also can be modeled as a two-person decision problem with a signaling information structure, only now the DMs form what will be described as a "team." To correspond more accurately to the Spence problem, the formulation will be modified to introduce the concept of "real-time information theory." This provides decision and control theorists with an understanding of information theory in their own terms. On the other hand, it provides information theorists with an entirely new way of looking at Shannon theory.

REFERENCES:

1. Y. C. Ho, I. Blau, and T. Basar, "A Tale of Four Information Structures," Lecture Notes in Math. Econ. and System Theory, Vol. 107, Springer Verlag, 1975.

2. Y. C. Ho and K. C. Chu, "Team Decision Theory and Information Structures in Optimal Control Problems--Part I," IEEE Trans. AC, Vol. AC-17, No. 1, Feb. 1972, pp. 15-22.

3. Y. C. Ho and K. C. Chu, "Information Structure in Many-Person Optimization Problems," Automatica, Vol. 10, 1974, pp. 341-351.

4. Y. C. Ho and F. K. Sun, "Value of Information in Zero-Sum Games," Math. Programming Studies, J. B. Wetts, ed., Vol. 5, 1976, pp. 211-226.

5. C. E. Shannon, "A Mathematical Theory of Communication," Bell Sys. Tech. J., Vol. 27, 1948, pp. 379, 623.

6. M. Spence, Market Signaling: Information Transfer in Hiring and Related Screening Processes, Cambridge, MA: Harvard University Press, 1974.

7. H. S. Witsenhausen, "A Counterexample in Stochastic Optimum Control," SIAM Journal Control, Vol. 6, No. 1, 1968, pp. 131-147.

8. H. S. Witsenhausen, "Separation of Estimation and Control for Discrete Time Systems," Proc. IEEE, Vol. 59, No. 11, Nov. 1971, pp. 1557-1566.

CHAPTER II

MARKET SIGNALING AND THE SPENCE MODEL

1. Introduction

In the job market model of Spence [1], [2], an employer must hire someone for a job without knowing how productive that individual will be. In other words, the employer has imperfect information about an individual's ability. Spence suggests that the employer can improve his information by looking on the job application for some signal, such as educational level. The employer offers wages based on the signal he sees; that is, a person with more education is offered higher wages, because the employer believes that the higher education indicates higher ability. The individual applying for the job, on the other hand, knowing he will receive wages based on his educational level, must decide how much education to get, taking into consideration that education is costly. When the employer's beliefs about the relationship between ability and education are confirmed by what the individuals actually do, then we have what Spence calls an equilibrium.

An interesting feature of this model is that there are multiple equilibrium solutions. In this chapter, we explain why this is true and prove new results about the Spence model. In order to do this, the model is formulated as a two-person nonzero-sum noncooperative decision problem with imperfect and dynamic information. The purpose of this is to clearly display the decision and control theoretic nature of the problem, in particular the role played by the dynamic information

structure. Under this formulation, new classes of multiple equilibria can be found and an explicit method for computing these new equilibria is given. Also, different properties of the solutions are investigated, such as stability and threshold effects.

2. Problem Statement

All potential employees will be considered together as decision maker one (DM1) and the employer as decision maker two (DM2). DM1's information is natural ability, denoted by the variable x. That is, each person makes a decision based on knowledge of his own true ability. This can be expressed as a mapping from ability to educational level, denoted by $\gamma_1(x) = u$, where u is the variable representing educational level. (We assume $x > 0$ and $u \geq 0$ to rule out the meaningless notions of "negative" ability and "negative" amounts of education.) The employer's information is the signal u, and his strategy is to offer wages as a function of education, denoted $\gamma_2(u) = v$, where v represents wages. We immediately see that this is the type of dynamic information structure defined in Chapter I as "signaling," where educational level u is the signal.

In Spence's model, signaling costs $c(u,x)$ and productivity $s(u,x)$ are functions of education level and ability.* Each individual applying for a job chooses the educational level to maximize his net profit, the difference between his wages and costs. For DM1, the entire employee

*Thus, educational level u not only serves as a signal about x, but also affects productivity directly when $s(u,x)$ is an explicit function of u.

population, the goal is to maximize the expected net profit, with expectation taken over the variable x. We assume that everyone, including the employer, knows the distribution of ability types throughout the population. Thus, the payoff function J_1 for the individuals is written

$$J_1(\gamma_1, \gamma_2) = E[\gamma_2(\gamma_1(x)) - c(\gamma_1(x), x)] \quad .$$

This is the same criterion that Spence proposes although he does not consider it in the context of a two-person decision problem.

Assuming utility units are appropriately defined, the employer would like to pay people no more than what they are worth; that is, he wants wages v to be less than or equal to productivity s. However, if v is strictly less than s, another employer could come along and offer wages greater than v but less than s, attract employees away from the first employer, and still make a profit. We will combine this idea of competition with the original proposal that wages not be greater than productivity in a single loss function for the employer by penalizing any deviation from s. Hence, the employer wants to choose a wage schedule γ_2 to minimize the quadratic loss function

$$J_2(\gamma_1, \gamma_2) = E[\gamma_2(\gamma_1(x)) - s(\gamma_1(x), x)]^2 \quad .$$

J_2 is a mathematical device to allow us to (1) reproduce Spence's result under our setup, and (2) focus on the equilibrium under competition without bringing in competition explicitly, thus avoiding the complication of a three-person decision problem. In Section 6 we will,

however, discuss the issue of competition directly, as was done in [3] and [4].

The problem is now in strategic form; i.e., the goal is to find the "optimal" strategies γ_1 and γ_2, where by optimality we mean finding the noncooperative Nash equilibrium, sometimes referred to as person-by-person optimality. This is defined as follows: (γ_1^*, γ_2^*) is a Nash equilibrium pair for the objective functions J_1 (maximize) and J_2 (minimize) if and only if

$$J_1(\gamma_1^*, \gamma_2^*) \geq J_1(\gamma_1, \gamma_2^*) \quad \forall \quad \text{admissible } \gamma_1$$

$$J_2(\gamma_1^*, \gamma_2^*) \leq J_2(\gamma_1^*, \gamma_2) \quad \forall \quad \text{admissible } \gamma_2 \quad .$$

That is, neither DM has the incentive to unilaterally deviate from the equilibrium solution. By standard manipulations [5], the first order necessary conditions for the Nash equilibrium are:

$$\max_u \; E_{/x}\,[\gamma_2(u) - c(u,x)] \;\Rightarrow\; \gamma_2' = \frac{\partial c}{\partial u} \triangleq c_u$$

$$\min_v \; E_{/u}\,[(v - s(u,x))^2] \;\Rightarrow\; \gamma_2(u) = v = E_{/u}\,(s)$$

where $'$ denotes d/du, and $E_{/x}(\cdot)$ denotes $E(\cdot\,|x)$.* It is clear that the second order sufficient conditions for the second equation hold, since J_2 is quadratic.

---

*That is, instead of solving for the strategies γ_1^* and γ_2^* in function space, we fix the arguments x and u and solve for the variables u and v, respectively.

The difficulty now is that $p(x|u)$, the underlying probability density function in the determination of $\gamma_2(u)$, is solution-dependent; that is, it cannot be evaluated until γ_1 is specified. A way out of this predicament is to guess that γ_1 is a one-to-one function. Then knowledge of u implies knowledge of x, so that $\gamma_2(u) = E_{/u}[s(u,x)] = s(u, x = \gamma_1^{-1}(u))$. Spence proves in [1] and [2] that the second order conditions for the first equation, namely, $\gamma_2'' - c_{uu} < 0$, are satisfied in this case under the assumptions

i) $c_u > 0$

ii) $c_{ux} < 0$

iii) $s_x > 0$.

A particular example from [2] in which $\gamma_2 = s$ is as follows:

EXAMPLE 2.1:

$$c = \frac{u}{x} \qquad\qquad s = x$$

Then, $\gamma_2' = 1/x$ and $\gamma_2 = x$, or

$$\gamma_2' = \frac{1}{\gamma_2} \quad .$$

This is a differential equation in γ_2 and has the one-parameter family of solutions

$$\gamma_2(u) = \sqrt{2u + 2k} \tag{2.1}$$

where k is the parameter. Since $\gamma_2 = x$,

$$\gamma_1(x) = u = \frac{1}{2}x^2 - k \quad . \tag{2.2}$$

Since $x > 0$, then γ_1 in (2.2) is, in fact, one-to-one, and our original assumption is verified. Equations (2.1) and (2.2) are the equilibrium solutions derived in [1] and [2], so that the two-person model recaptures Spence's original results.

An important property of the solutions is that varying k produces a continuum of multiple equilibria. Since

$$\frac{dJ_1^*(k)}{dk} = E\left[\frac{1}{x}\right] > 0$$

(where $J_1^*(k) = J_1(\gamma_1(k), \gamma_2(k))$, γ_1 and γ_2 as in (2.1) and (2.2)), the equilibria parameterized by larger k give larger expected net profit to DM1 than those with smaller k. For DM2, varying k does not matter, since γ_2 always equals x and J_2 remains zero. When one equilibrium solution is better than another solution for at least one DM without harming the other DM, the former solution is called "Pareto-superior" to the latter. Thus, solutions (2.2) with larger k are Pareto-superior to those with smaller k.

Spence also works out an example where x and u are discrete random variables [2]:

EXAMPLE 2.2: Let $x \in \{1, 2\}$. Let

q = fraction of population of type $x = 1$

$1-q$ = fraction of population of type $x = 2$

$c = u/x \qquad s = x$.

Suppose the employer guesses a relationship between ability and education that results in the following conditional density function and wage schedule:

$$\Pr\{x=1 \mid 0 \le u < u^*\} = 1 \quad \Rightarrow \quad \gamma_2(u) = 1 \quad \text{for} \quad u < u^*$$

$$\Pr\{x=2 \mid u \ge u^*\} = 1 \quad \Rightarrow \quad \gamma_2(u) = 2 \quad \text{for} \quad u \ge u^* \quad .$$

γ_2 is a two-level step function as shown in Figure 2.1. Since the cost of education is a monotonically increasing function of u for fixed x, the net profit $\gamma_2 - c$ will be maximized only at the education levels $u = 0$ or u^*. Thus, the individuals will choose either $u = 0$ or u^*. For $x = 1$,

$$\max\{\gamma_2 - c\} = \max\{1, 2 - u^*\} \quad .$$

For $x = 2$,

$$\max\{\gamma_2 - c\} = \max\left\{1, 2 - \frac{u^*}{2}\right\} \quad .$$

Therefore, in order to have consistency with the employer's beliefs, we must have:

$$\left.\begin{aligned} \gamma_1(x=1) = 0 &\iff 1 > 2 - u^*, \quad \text{or} \quad u^* > 1 \\ \gamma_1(x=2) = u^* &\iff 2 - \frac{u^*}{2} > 1, \quad \text{or} \quad u^* < 2 \end{aligned}\right\} \iff 1 < u^* < 2 \qquad (2.3)$$

Inequality (2.3) is the equilibrium condition for this discrete example. Varying the parameter u^* between 1 and 2 again results in a continuum of multiple equilibria. Also,

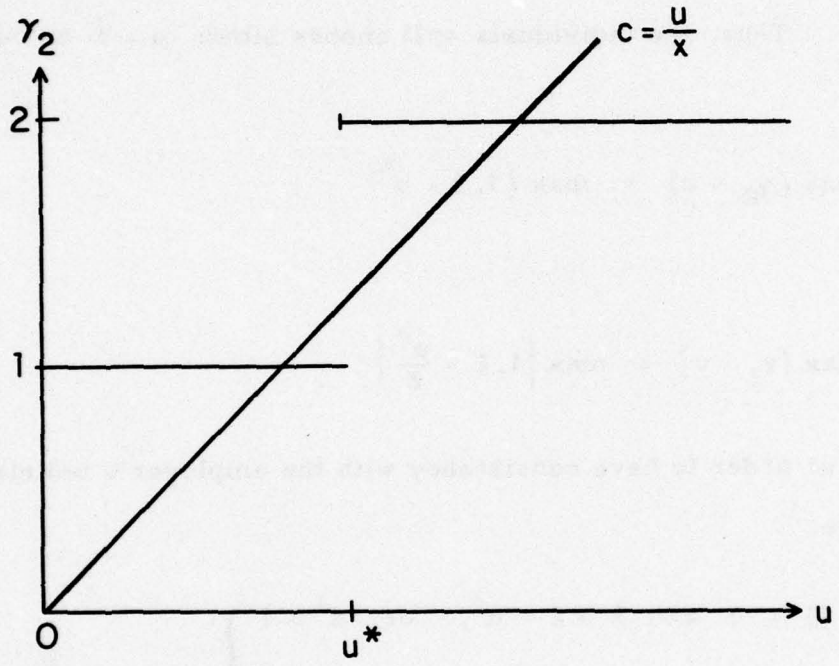


FIG. 2.1 WAGE SCHEDULE FOR EXAMPLE 2.2

$$J_1^*(u^*) = 1 \cdot q + \left(2 - \frac{u^*}{2}\right)(1 - q)$$

$$J_2^*(u^*) = 0 \quad .$$

As u^* decreases, J_1^* increases. Therefore, solutions with smaller u^* are Pareto-superior to those with larger u^*.

Because the information structure is dynamic, the employee has complete control on what the employer can infer from his observation u about the underlying state of nature x. Thus, it is not too surprising that by allowing the employer to make different kinds of inference on the functional relationship between x and u, different functional forms for the equilibria can be obtained.

3. Comparison of Equilibria

In Example 2.1, γ_1 is a one-to-one mapping from a continuous set of abilities to a continuous set of educational levels. In Example 2.2, γ_1 is also one-to-one, but this time the sets are discrete. In both of these examples, the employer can precisely determine ability from merely looking at the signal. In our model, we will obtain equilibria somewhere between these; our equilibria involve a continuous range of abilities but a discrete, finite number of given signals. Thus, our mappings from ability to signals are many-to-one.* We believe our equilibria are intuitively appealing for several reasons. First of all,

*Our equilibria are actually many-to-one solutions for Example 2.1. The available range of signals remains continuous, but only a finite number of signals are actually chosen by the employees.

in actuality, there are only a discrete number of educational levels at which wages are offered, for example, bachelor, master, and doctorate degrees.* Secondly, many different types of people choose the same signal, suggesting a *many-to-one mapping*. Lastly, employers are limited in the amount of information processing they can do, so that they can handle only a discrete number of signals.

Another property Spence's discrete example has in common with his continuous one is that there are multiple equilibria, in fact, a continuum of multiple equilibria. As mentioned earlier, some are Pareto-superior to others in the sense that they give a higher expected net profit J_1 to DM1. Spence points out that the Pareto-inferior solutions are inefficient in the sense that people are overinvesting in the signal by purchasing more education that is necessary to signal their ability levels. Spence [3] and Riley [4] have discussed how to choose the Pareto-optimal solution, if possible, that is, the solution which has no other solutions Pareto-superior to it, in order to eliminate or reduce the inefficiencies. However, they assume that the employer has the power to manipulate both the wage schedule γ_2 and the signal levels u by changing the parameters of the problem, in the first example by varying k, and in the second by varying u^*. In effect, this is equivalent to changing the signals already existing in the market. In our equilibria, we assume that the parameters, and hence the signals, are fixed exogenously. This reduces the multiple equilibria

*Although "pseudo" educational levels, such as "master's degree with two years experience," have been created over time, they are still discrete.

to a single equilibrium, in general. The justification for this assumption is that when the employer comes into the market, the educational levels used for pay scales are already determined. Only after a long period of time can new levels be established. Table 2.1 summarizes the differences between our solutions and Spence's.

TABLE 2.1

Comparison of Equilibrium Classes

	Ability	Signal (e.g. Education)
Spence	Discrete	Discrete-not fixed
Spence-Riley	Continuous	Continuous-not fixed
Ho-Kastner	Continuous	Discrete-fixed

4. New Multiple Equilibrium Classes

In Example 2.1, we began by guessing γ_1 was one-to-one, determined γ_2 from this γ_1, and then found that the resulting solution was consistent with the original guess. In the second example, we guessed γ_2 as a function of a parameter and then through γ_1 determined the values of the parameter that would give consistency. Thus, as mentioned in the introduction, an equilibrium can be described as the solution to an implicit equation resulting from a mathematically self-consistent loop, as shown in Figure 2.2, where $p(x|u)$ is the

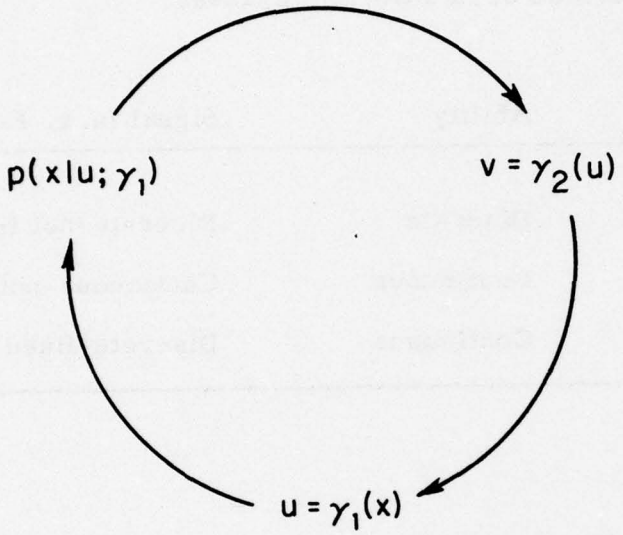


FIG. 2.2 SELF-CONSISTENCY LOOP ILLUSTRATING IMPLICIT EQUATION

conditional probability density of x given the signal u, which is determined by γ_1. To find the new equilibrium classes, we guess γ_1 in terms of some parameters and determine the conditions on the parameters so that the resulting γ_1 from circling the loop is the original guess γ_1.

Figures 2.3 and 2.4 show the kind of equilibria we are looking for. We assume ability x lies in some fixed range $[x_0, x_N]$. Let $x_1, x_2, \ldots, x_{N-1}$ be points inside the interval such that

$$x_0 < x_1 < x_2 \cdots < x_{N-1} < x_N \quad . \tag{2.4}$$

Let $u \in [u_0, u_N]$, and assume that ability types within each subinterval $[x_i, x_{i+1})$ choose the same signal u_i. The endpoints x_i of these subintervals, except for x_0 and x_N, will be called "breakpoints". No single person chooses the breakpoints; they just reflect how the entire employee population divides itself. More precisely, DM1's strategy is as follows (see Figure 2.3):

$$\gamma_1(x) = \begin{cases} u_i, & x \in [x_i, x_{i+1}) \ \forall \ i = 0, \ldots, N-2 \\ u_{N-1}, & x \in [x_{N-1}, x_N] \end{cases} .$$

It is clear from J_1 that we must have $u_0 < u_1 < \cdots < u_{N-1}$, for otherwise, wages ($\gamma_2$) would be a decreasing function of education level, an intuitively unappealing result.* Since γ_1 assumes discrete

*We are assuming here that γ_2 is a monotonically increasing function of educational level.

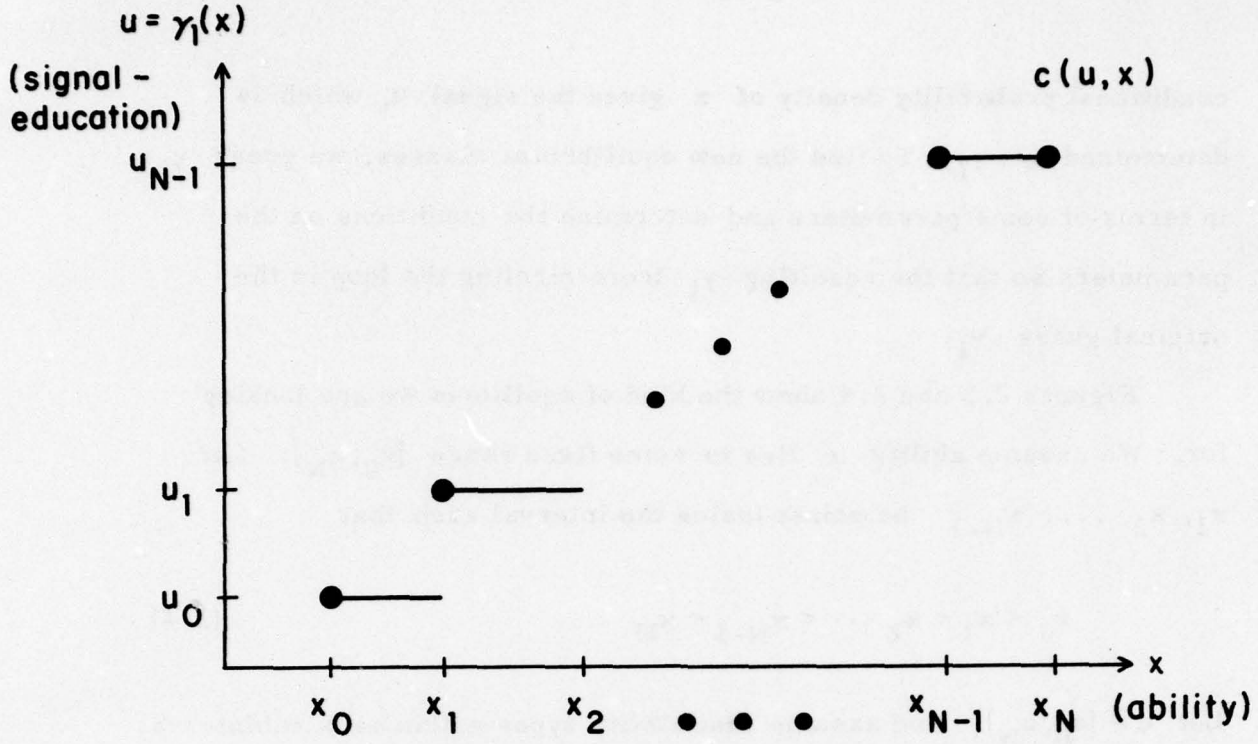


FIG. 2.3 STRATEGY FOR EMPLOYEE POPULATION

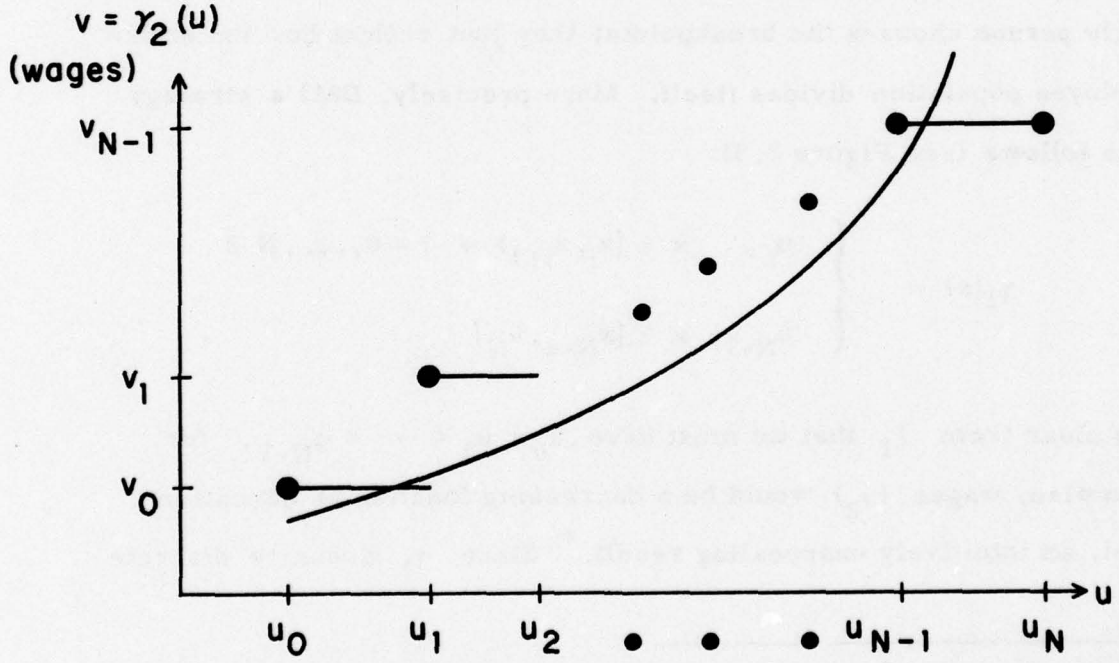


FIG. 2.4 EMPLOYER'S STRATEGY

values over a finite number of intervals, the assumptions for this problem will be a discrete version of the first two assumptions above, that is, i') $\Delta c/\Delta u > 0$ and ii') $\Delta^2 c/\Delta x \Delta u < 0$. The third assumption will not be needed to prove the sufficient conditions for this class of solutions.

The employer can now compute his strategy, after he computes the conditional density function (as shown in Figure 2.2):

$\forall\ i = 0, \ldots, N-1$,

$$p(x|u_i) = \begin{cases} \dfrac{p(x)}{F(x_{i+1}) - F(x_i)}, & x \in [x_i, x_{i+1}) \ ([x_{N-1}, x_N] \text{ for } i = N-1) \\ 0 & \text{otherwise} \end{cases}$$

(p is the probability density function and F is the distribution function) and

$$\gamma_2(u_i) = E_{/u_i}[s(u_i, x)] = \frac{\int_{x_i}^{x_{i+1}} s(u_i, x)p(x)\ dx}{F(x_{i+1}) - F(x_i)}$$

$$\triangleq v_i$$

$$\triangleq g_i(x_i, x_{i+1}) \qquad . \qquad (2.5)$$

The variables v_i represent the actual wage values, and the functions g_i show the dependence of wages on x_i and x_{i+1}. So far, γ_2 is defined only for the discrete signals u_i, $i = 0, \ldots, N-1$. In order to have our equilibria be many-to-one solutions for Example 2.1, γ_2 must be a Nash solution in the strategy space of measurable mappings

defined over the entire interval $[u_0, u_N]$. We arbitrarily define γ_2 completely as

$$\gamma_2(u) = \begin{cases} v_i, & u \in [u_i, u_{i+1}), \quad i = 0, \ldots, N-2 \\ v_{N-1}, & u \in [u_{N-1}, u_N] \end{cases} .$$

Thus, the employer's strategy also looks like a step function (Figure 2.4). As will now be shown, this particular form for γ_2 gives us the results we need for a simple equilibrium condition.

Given the wage schedule γ_2, DM1 can continue around the loop and compute a new strategy $\hat{\gamma}_1$:

$$\hat{\gamma}_1(x) = \arg \max_u [\gamma_2(u) - c(u,x)] .$$

As shown in Figure 2.4, assumption i') that $c_u > 0$ implies that DM1 will only consider choosing among $u_0, \ldots, u_{N-1}$. Therefore,

$$\hat{\gamma}_1(x) = \arg \max_{u_i \in \{u_0, \ldots, u_{N-1}\}} [g_i(x_i, x_{i+1}) - c(u_i, x)] .$$

In order to attain self-consistency and have $\hat{\gamma}_1 \equiv \gamma_1$, we want, for all $i = 0, \ldots, N-1$ (omitting the arguments of g_i for simplicity)

$$g_i - c(u_i, x) > g_j - c(u_j, x) \quad \forall \; j \neq i \quad \text{and} \quad x \in [x_i, x_{i+1}) .$$

The following proposition states that if people whose ability levels are at a breakpoint x_i are indifferent between the educational levels u_i and u_{i-1}, then $\hat{\gamma}_1 \equiv \gamma_1$ for all x except the breakpoints.

PROPOSITION 2.1: If

$$g_i - c(u_i, x_i) = g_{i-1} - c(u_{i-1}, x_i) \quad \forall \; i = 1, \ldots, N-1 \tag{2.6}$$

then

$$g_i - c(u_i, x) > g_j - c(u_j, x) \tag{2.7}$$

for all $j \neq i$ and for all $x \in (x_i, x_{i+1})$.

Proof. From assumptions i') and ii'),

$$c(u_i, x) - c(u_{i+1}, x) < c(u_i, x_{i+1}) - c(u_{i+1}, x_{i+1}) \quad \forall \; x < x_{i+1} \; .$$

From (2.6),

$$g_i = g_{i+1} - c(u_{i+1}, x_{i+1}) + c(u_i, x_{i+1}) \; .$$

Then

$$g_i - c(u_i, x) > g_{i+1} - c(u_{i+1}, x) \quad \forall \; x < x_{i+1} \; .$$

For all $x \in (x_i, x_{i+1})$,

$$g_i - c(u_i, x) > g_{i+1} - c(u_{i+1}, x) > g_{i+1} - c(u_{i+2}, x) \ldots$$

so that

$$g_i - c(u_i, x) > g_j - c(u_j, x) \quad \forall \; j > i \; .$$

Similarly,

$$g_i - c(u_i, x) > g_{i-1} - c(u_{i-1}, x) \quad \forall \; x > x_i$$

implies

$$g_i - c(u_i, x) > g_j - c(u_j, x) \quad \forall \; j < i, \quad \forall \; x \in (x_i, x_{i+1}) \; . \qquad \text{Q. E. D.}$$

The following corollary states that the indifference condition (2.6) implies that at the breakpoint x_i, u_i and u_{i-1} are preferred over all other signals.

COROLLARY 2.1. Given (2.6),

$$g_i - c(u_i, x_i) > g_j - c(u_j, x_i) \quad \forall \; j \neq i-1, i, \qquad i = 1, \ldots, N-1 \tag{2.8}$$

Proof. Suppose there exists $j \neq i - 1, i$ such that $g_j - c(u_j, x_i) > g_i - c(u_i, x_i)$. If $j > i$, then by assumptions i') and ii'),

$$-c(u_i, x_i) + c(u_j, x_i) > -c(u_i, x) + c(u_j, x)$$

so that

$$g_j - c(u_j, x) > g_i - c(u_i, x) \quad \forall \; x \in (x_i, x_{i+1}) \; ,$$

which contradicts (2.7). Similarly, if $j < i - 1$, then

$$g_j - c(u_j, x) > g_{i-1} - c(u_{i-1}, x) \quad \forall \; x \in (x_{i-1}, x_i)$$

which also contradicts (2.7). Q. E. D.

Therefore, if (2.6) holds, and if we define $\hat{\gamma}_1$ at the breakpoints as

$$\hat{\gamma}_1(x_i) = u_i, \quad i = 0, \ldots, N-1 \quad \text{and} \quad \hat{\gamma}_1(x_N) = u_{N-1} \tag{2.9}$$

then $\hat{\gamma}_1 \equiv \gamma_1$ for all $x \in [x_0, x_N]$.

Thus, for fixed levels of education (u), "optimizing" means choosing the breakpoints $\{x_i\}$, what the employee population as a whole should do, and wage levels $\{v_i\}$, what the employer should do.* We have shown that the necessary and sufficient conditions for optimality reduce to the set of equalities (2.6) and inequalities (2.4) involving the x_i's and v_i's.** The inequalities say that the breakpoints should be "in order". The equalities say that if the people whose ability level is a breakpoint, say x_1, are indifferent between choosing educational level u_0 and receiving wage v_0, and choosing u_1 and receiving v_1, then the system is in equilibrium and people are paid equal to the expected productivity of their particular signaling group. We have reduced the Nash equilibrium of a two-person decision problem to a feasible solution of equalities and inequalities. Equations (2.6) provide an explicit method for computing the equilibria. If the u_i's are varied or if the number of signals N is changed, then there are multiple equilibria. But, if as mentioned earlier, the signals are fixed, then there are, in general, no multiple solutions.

---

*This problem is different from those of Section 2. In Examples 2.1 and 2.2, u was found for each individual x. Here, the entire employee population is considered in determining where the breakpoints should be, and thus, what signals should be chosen.

**These conditions are clearly sufficient for optimality. However, they are necessary for optimality only in the class of solutions we have guessed, namely, many-to-one in the manner of Figures 2.3 and 2.4.

EXAMPLE 2.3: Let

$$c(u,x) = u/x \qquad p(x) \text{ uniform over } [x_0, x_N]$$

$$s(u,x) = x \qquad N = 3$$

From the definition of γ_2 at equilibrium given by (2.5),

$$v_i = \frac{x_i + x_{i+1}}{2} , \qquad i = 0, 1, 2 \quad .$$

From (2.6)

$$x_i = \frac{u_i - u_{i-1}}{v_i - v_{i-1}} , \qquad i = 1, 2 \quad .$$

Combining these, we have the equilibrium conditions

$$x_1 = \frac{2(u_1 - u_0)}{x_2 - x_0} \tag{2.10a}$$

$$x_2 = \frac{2(u_2 - u_1)}{x_3 - x_1} \quad . \tag{2.10b}$$

Conditions (2.10) depend on u only through the differences $u_1 - u_0$ and $u_2 - u_1$, since c is linear in u and s is independent of u.

If $x_0 = 1$, $x_3 = 2.5$, $u_1 - u_0 = u_2 - u_1 = 1$, then the pair (x_1, x_2) satisfying (2.10) and (2.4) is*

*The pair (-3.1, .36) also satisfies (2.10) but not (2.4). In every example we tried, only one of the pairs satisfying (2.6) also satisfied (2.4), but we have not ruled out the possibility that both pairs might be solutions.

$$\left(\frac{1}{4}\left(\sqrt{89} - 3\right),\ \frac{1}{10}\left(\sqrt{89} + 13\right)\right) \approx (1.6,\ 2.24) \quad .$$

The title of this section promises multiple equilibria, but here we have a unique equilibrium. This occurs because, as mentioned above, we assume u_0, u_1, and u_2 are fixed.

Table 2.2 gives numerical results for the cases $N = 2, 3, 4$. Figures 2.5 and 2.6 illustrate that for $N = 4$, γ_1 and γ_2 are already beginning to look like the square and square root functions, respectively, which are the solutions to Example 2.1, Spence's continuous one-to-one case.

Other functions of c and s, and other probability densities $p(x)$, such as the Gaussian distribution, also produce new classes of multiple equilibria for different values of N, but the details are omitted here.

TABLE 2.2

Numerical Examples for the Uniform Distribution

N	$x_0, x_1, \ldots$	$u_0, u_1, \ldots$	$v_0, v_1 \ldots$
2	1, 1.33, 2.5	0, 1	1.17, 1.92
3	1, 1.6, 2.24, 2.5	0, 1, 2	1.3, 1.92, 2.37
4	1, 1.6, 2.23, 2.6, 3	0, 1, 2, 3	1.3, 1.92, 2.42, 2.8

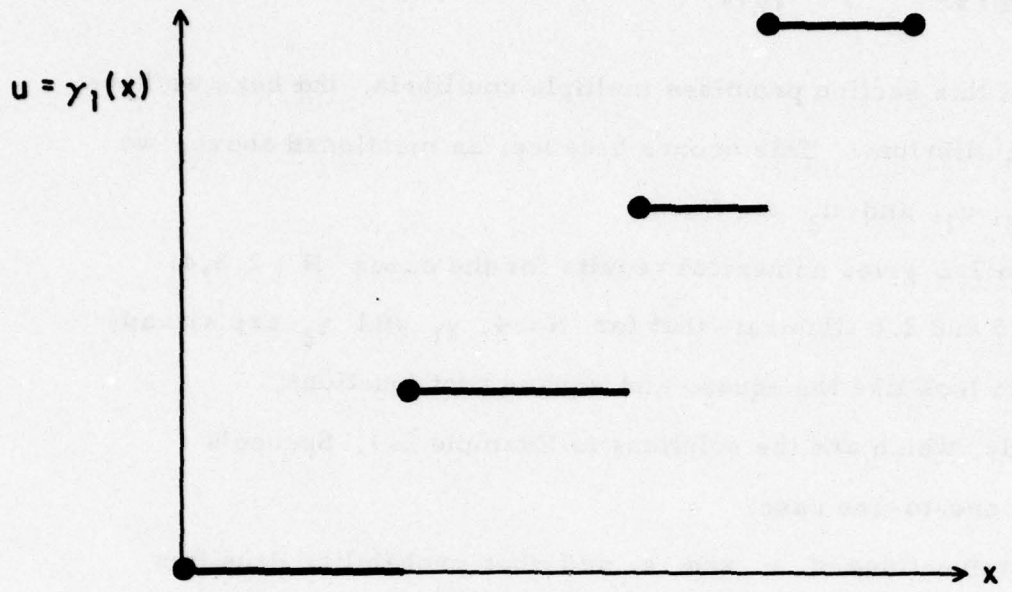


FIG. 2.5 EQUILIBRIUM EDUCATION LEVELS FOR N=4

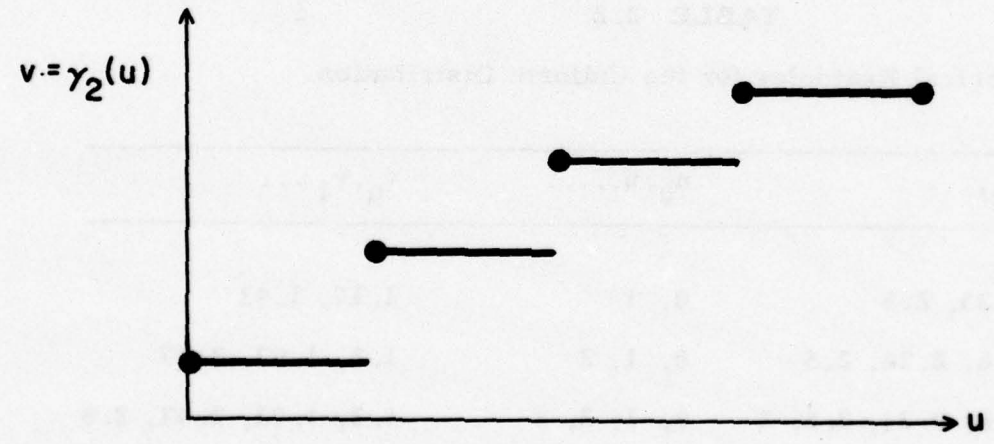


FIG. 2.6 EQUILIBRIUM SCHEDULE FOR N=4

5. Adjustment Procedure and Stability

An important question to ask about these new equilibrium classes is whether they are stable. That is, if the system is not in equilibrium, will it return to equilibrium. In order to answer this question, an adjustment procedure must be outlined for each decision maker, describing how he would react if the system were in disequilibrium. Then we can see if these actions bring all of the DMs back to the equilibrium.

A reasonable adjustment scheme for the employer is the one Spence proposes in his definition of equilibrium. He states that the employer always pays wages equal to the expected productivity, based on the statistical data revealed by the previous employee population. The data the employer observes after the employees are hired are just the breakpoints, that is, which range of abilities choose which signal. Thus, he uses the data he observed in the past stage to make his estimate in the current stage. This is written

$$v_i(t) = g_i(x_i(t), x_{i+1}(t)) = E_{/u_i}[s(u_i, x)]$$

$$= \frac{\int_{x_i(t)}^{x_{i+1}(t)} s(u_i, x)p(x)\, dx}{F(x_{i+1}(t)) - F(x_i(t))} \quad . \tag{2.11}$$

We call this "full equilibrium adjustment", because (2.11) is just the equilibrium condition (2.5). That is, the objective function J_2 is minimized at each stage.

The adjustments that the employee population makes to the change in wage schedules can be argued, on the other hand, to be more gradual or infinitesimal. After the employer has adjusted his wage schedule, the individuals at the breakpoints, who were once indifferent, now have a clear choice as to which signal they prefer. This is reflected in the shifting of the breakpoints. People on one side of the breakpoint slowly drift over to the other side as they learn how to respond to the wage schedule. The net result can be modeled by a set of steepest ascent equations for the breakpoints,

$$\dot{x}_i(t) = \epsilon \frac{\partial J_1}{\partial x_i} \tag{2.12}$$

where $\epsilon > 0$ is a constant defining the infinitesimal incremental step. In other words, no single individual changes x_i; the shift is due to the combined action of the entire employee population. We call (2.12) "partial equilibrium adjustment" because, although each step moves in the direction of maximizing J_1, J_1 is not actually maximized at each stage. This defines the other half of the adjustment procedure.

Substituting (2.11) into (2.12) results in a set of differential equations

$$\dot{x}_i \triangleq \delta_i(x_1, \ldots, x_{N-1}), \quad i = 1, \ldots, N-1 \quad .$$

This has reduced the problem of adjustment of individual actions to the question of stability of a set of differential equations. The stability result we need is a version of a Lyapunov-type stability theorem due to Malishevskii [6]. His study of stability of individual actions in

goal-oriented behavior is in the same spirit as our problem. His theorem, restated in general terms, is as follows:

THEOREM (Lyapunov-Malishevskii): Let $\dot{x}_i = \delta_i(x)$, where $x = (x_1, \ldots, x_n)^T$ is an element in some domain $D \subset R^n$. Define the matrix

$$A = \left(\frac{\partial \delta_i}{\partial x_j} \right) \quad .$$

If $A + A^T < 0$ for all x in D and if the equilibrium point x^* (where x^* satisfies $\delta_i(x^*) = 0$ for all i) exists in D, then any trajectory $x(t)$ which remains in D converges to x^* (uniform asymptotic global stability).

The following example is an illustration of how this theorem can be applied to the job market model. From the definition of J_1,

$$J_1 = E[\gamma_2 - c] = \sum_{k=0}^{N-1} \int_{x_k}^{x_{k+1}} [v_k - c(u_k, x)]\, p(x)\, dx$$

and

$$\frac{\partial J_1}{\partial x_i} = [v_{i-1} - c(u_{i-1}, x_i) - (v_i - c(u_i, x_i))]\, p(x_i) \quad .$$

EXAMPLE: Consider Example 2.3 above, and let ϵ' represent $\epsilon p(x_i) = \epsilon/(x_N - x_0)$. Then

$$v_i = \frac{x_i + x_{i+1}}{2}$$

$$\dot{x}_1 = \epsilon' \left(\frac{1}{x_1} + \frac{1-x_2}{2} \right) \triangleq \delta_1$$

$$\dot{x}_2 = \epsilon' \left(\frac{1}{x_2} + \frac{x_1-2.5}{2} \right) \triangleq \delta_2$$

Let $D = \{(x_1, x_2) : x_1 > 0,\ x_2 > 0\}$. By inspection, it is clear that the trajectory $x(t) = (x_1(t),\ x_2(t))$ defined by δ_1 and δ_2 remains in D.

$$A + A^T = \epsilon' \begin{bmatrix} -\frac{2}{x_1^2} & 0 \\ 0 & -\frac{2}{x_2^2} \end{bmatrix} < 0 \quad \text{in } D .$$

Therefore, in D, $x(t) \to x^* \cong (1.6,\ 2.24)$, independent of ϵ'. This is as it should be, since one cannot in general be certain of the value of ϵ. The above procedure, in fact, also holds for arbitrary discrete levels of signals; that is, it is independent of N.

This is not the whole story, however, because the breakpoints still must satisfy (2.4), that is, be "in order". The previous example can serve to describe what might happen before the equilibrium is reached. The stability result says that x(t) will converge to (x_1^*, x_2^*). But the extra constraint of order defines a region where $x_0 < x_1 < x_2 < x_3$ which we call the "feasible region" (FR), as shown in Figure 2.7. Even if a trajectory starts inside this region, it may leave the region before it reaches the equilibrium point, as illustrated by the dotted curve in Figure 2.7. If this happens, it means that two breakpoints, or a breakpoint and an endpoint, have coalesced. One of

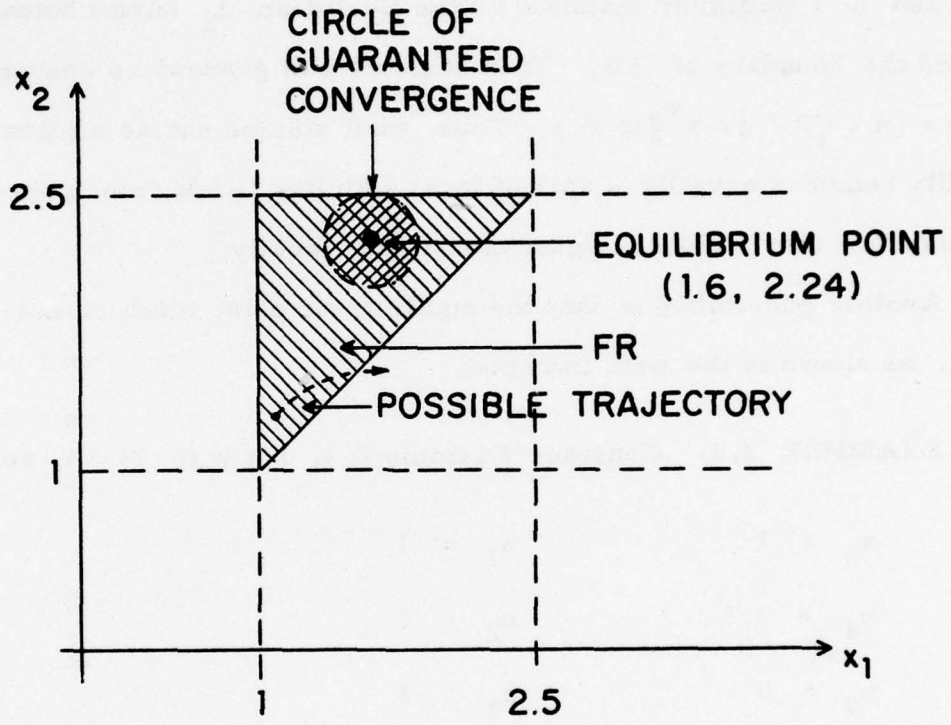


FIG. 2.7 STABILITY FOR EXAMPLE 2.3

the breakpoints has disappeared, meaning that one of the signals is no longer being chosen by any individuals. The system then drops back down to the next lower number of signals. However, there is always a circle of initial points where convergence is guaranteed, because Malishevskii proves that the norm of the vector $x(t)$ monotonically decreases. This circle is defined as follows:

Let d^* = minimum distance (in the Euclidean ℓ_2 norm) between x^* and the boundary of FR. Then the circle of guaranteed convergence = $\{x \in FR : \|x-x^*\| \leq d^*\}$. Thus, what started out as a global stability result is actually a sort of local stability result, since convergence for this problem is guaranteed only locally.

Another possibility is that the equilibrium point itself is not in FR, as shown in the next example.

EXAMPLE 2.4: Consider Example 2.3, but with $N = 4$, and

$$x_0 = 1 \qquad u_1 = 1$$

$$x_4 = 2.5 \qquad u_2 = 2$$

$$u_0 = 0 \qquad u_3 = 3$$

Then the solution to the equilibrium equality conditions with all positive components

$$(x_1^*, x_2^*, x_3^*) = (2.2, 1.9, 3.3)$$

does not satisfy the "order" condition (2.4). This phenomenon does not depend on N, the number of levels. Table 2.2, Figure 2.5 and

Figure 2.6 demonstrate an example where the equilibrium breakpoints for $N = 4$ do lie inside the feasible region.

6. Competition and Specialization

6.1 Specialization

So far, we have assumed that the employer hires people of all abilities in $[x_0, x_N]$. Spence [3] and Riley [4] show how specialization; i.e., hiring people of only some abilities, can lead to nonexistence of an equilibrium when competition from other employers is brought in explicitly. In particular, Spence states that with specialization (1) a one-step wage schedule ($N = 1$ in our notation) definitely cannot be an equilibrium, and (2) if this one-step schedule is preferred by all employees to multistep (or continuous) schedules, then there is no equilibrium in the market. We will show that this last conclusion also holds for our equilibrium classes. However, we will also show in Section 6.2 that the nonexistence of an equilibrium can be partially resolved through use of the criterion function J_2.

The argument for (1) proceeds by way of Example 2.3 summarized as the $N = 3$ case in Table 2.2. The only available signals are $u = 0, 1, 2$. Suppose an employer ignores the last two signals and offers the one-step schedule

$$\gamma_2(u) = E[x] = \frac{1+2.5}{2} = 1.75 \quad \forall\, u \geq 0$$

as shown in Figure 2.8. Let

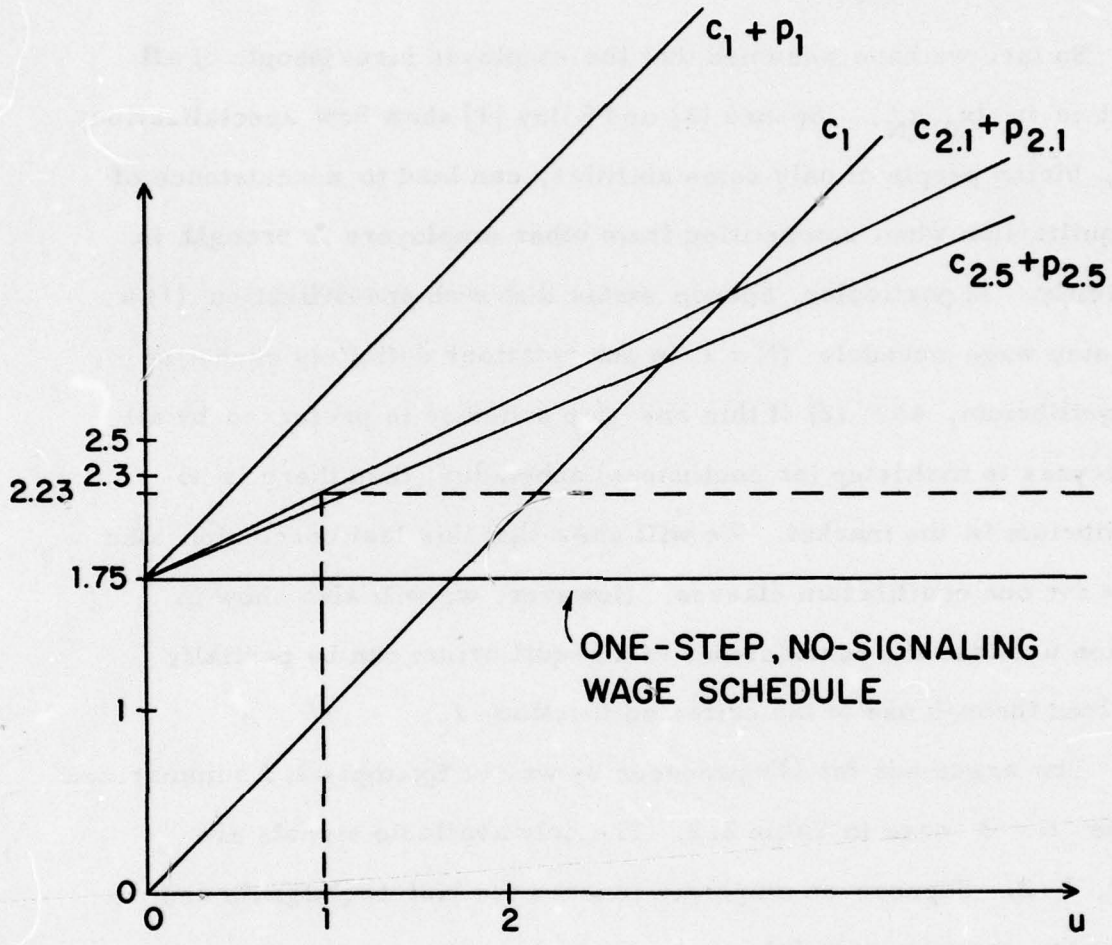


FIG. 2.8 SPECIALIZATION

p_x = optimal net profit of individual of type x

$= \gamma_2(u^*) - c(u^*, x)$

where

$$u^* = \arg\max_u [\gamma_2(u) - c(u, x)] \quad .$$

To express $c(u, x)$ as a function of u for fixed x, we use instead the notation $c_x(u)$, where

$$c_x(u) = \frac{u}{x}, \quad \text{for this example.}$$

If $\gamma_2(u) = 1.75$ for all u, then $u^* = 0$ and $p_x = 1.75$ for all x. Figure 2.8 shows the cost curve (line, in this case) c_1 and the cost curve shifted by the optimal net profit, $c_1 + p_1$. The line $c_x + p_x$ is the indifference line for a person of ability x; that is, any wage offered along this line gives net profit $p_x = 1.75$. Thus, any wage in the region above the line is preferred to the original one-step schedule since the net profit is greater than p_x.

Suppose an employer who was able to specialize offered the schedule in Figure 2.8 shown by the dotted line; that is,

$$\gamma_2(u) = \begin{cases} 0 \ , & 0 \le u < 1 \\ 2.23 \ , & u \ge 1 \end{cases} \quad .$$

Then, by the previous argument, all persons of ability types $x \in [2.1, 2.5]$ will prefer this new schedule. Since their average productivity is 2.3, which is greater than the wage 2.23, the employer makes a profit. The original one-step equilibrium is destroyed, and

Spence's first statement is demonstrated for this example. Whereas Spence demonstrated this by allowing the employer to change the discrete signaling levels u, we have demonstrated this without creating new signals, just ignoring some of the already existing signals.

To demonstrate the second statement, we must show how it can happen that all individuals prefer the one-step (N = 1) wage schedule. For this example, since there are three signals, there are three possible equilibrium wage schedules, corresponding to N = 1, N = 2, and N = 3. Figure 2.9 shows the N = 3 and N = 1 wage schedules. Wages in region R are preferred over the N = 3 schedule by those people whose abilities are at the endpoints of the ability interval, namely x = 1 and 2.5. The following argument shows that this region is also preferred by all abilities inbetween as well. All other shifted cost lines are also indifference lines and so must, by construction, pass through one of the points A, B, or C in Figure 2.9 with slopes between that of $c_1 + p_1$ and $c_{2.5} + p_{2.5}$, namely, 1 and 2/5. Thus, none of these cost lines will intersect the region R, so that the one-step wage schedule is preferred by everyone to the N = 3 schedule. Thus, everyone chooses $u_0 = 0$ and signaling ceases. Graphically, we see that this is true when the intersection of the expected productivity line given by $v = E(x) = (x_0 + x_N)/2$ with the v-axis lies above that of the line $c_{x_N} + p_{x_N}$. More precisely, this condition can be stated as:

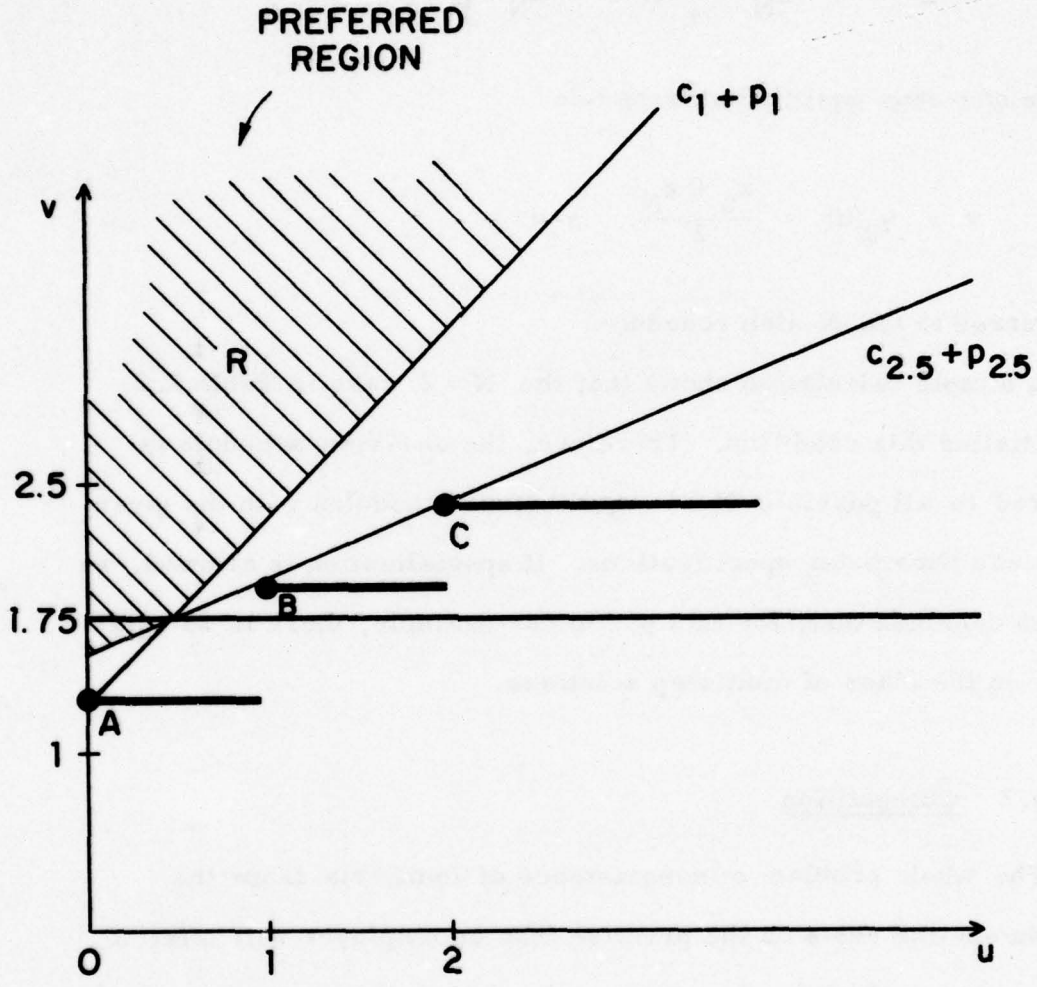


FIG. 2.9 ONE-STEP VS. THREE-STEP WAGE SCHEDULE

PROPOSITION 2.2. Consider the equilibrium N-step schedule of Example 2.3. If

$$\frac{x_0 + x_N}{2} > \frac{u_0}{x_N} + \left[v_{N-1} - \frac{u_{N-1}}{x_N} \right],$$

then the one-step equilibrium schedule

$$v = \gamma_2(u) = \frac{x_0 + x_N}{2} \quad \forall\, u$$

is preferred to the N-step schedule.

A simple calculation shows that the $N = 2$ case in Table 2.2 also satisfies this condition. Therefore, the one-step schedule is preferred to all possible $N > 1$ equilibrium schedules with the given signals and parameter specifications. If specialization is allowed, we can then conclude that, for this particular example, there is no equilibrium in the class of multistep solutions.

6.2 Competition

The whole problem of nonexistence of equilibria from the previous section rests on the premise that an employer will offer a one-step wage schedule when it is preferable to the employees, in order to compete with other employers. However, in analyzing this competition, we have considered J_1, but have totally neglected J_2. The outcome is the rather nonintuitive result that people of the highest ability would sometimes prefer to be paid the same as people of the lowest ability. This, it would seem, would lead to much job

dissatisfaction. Of course, if the number of signaling levels were fixed, or there were no other employers, then the individual would have no choice but to maximize J_1 over the available signals. But in the situation described above, neither of these is the case. The actual signals are exogenously given but the number of signals to be used is not. A persuasive case might be made that an employer offering a more differentiated wage schedule (e.g., N = 3) might very well be preferred by the employee population than one who offers the one-step schedule, even though the latter schedule makes more pure economic sense to the employees from the viewpoint of J_1. However, people are also concerned with being paid nearer to what they are worth. We submit that our J_2, the mean square criterion, is an attempt to capture this effect. To justify this conclusion in our decision-theoretic framework, we must show that the value of J_2 for the one-step schedule is larger than the value for the multi-step schedule; that is, the one-step schedule is less preferable to the employer by being less competitive. This is, in fact, the case for the example in Figure 2.9 where $J_2(N = 1) = .1875$ and $J_2(N = 3) = .0278$.* In other words, if additional signals are available, then there is, in general, an incentive for an employer to offer a finer schedule when it leads to a better J_2. A logical consequence of this argument is that the

*A simple calculation shows that for $c = u/x$, $s = x$, and $p(x)$ uniform

$$J_2 = \frac{1}{12(x_N - x_0)} \sum_{i=0}^{N-1} (x_{i+1} - x_i)^3 .$$

employer would prefer more and more signals to differentiate people of different abilities until Spence's continuous equilibrium is reached (as in Example 2.1), where every ability level is paid its productivity, and $J_2 = E[(\gamma_2 - s)^2] = 0$.

If this is true, then why don't we see employers constantly creating new signals in hopes of attaining the continuous one-to-one equilibrium. First of all, we argued above that it is very difficult to create new educational levels, and assumed that the signals were fixed exogenously when an employer entered the market. Secondly, adding new signals does not <u>necessarily</u> improve J_2, since we have the constraints that the old signals cannot be easily discarded and the breakpoints must be in order. For example, if the new signal $u_3 = 2.25$ is introduced for the case $N = 3$ in Figure 2.9, and if the employees and employers adjust so as to settle down at a new equilibrium at the $N = 4$ level (see Section 5 on adjustment and stability), then $J_2(N = 4) = .0355 > .0278 = J_2(N = 3)$. In fact, it can be shown that any $u_3 > 2$ which produces an equilibrium solution in the feasible region yields a J_2 value greater than $J_2(N = 3)$. Thirdly, for each new signal there are attendant costs of transmission and administration. In a sufficiently differentiated wage schedule, these second order costs must be accounted for and traded against the advantages of new signals. Consequently, we do not see the constant creation of new signals in the short run nor the eventual infinite differentiation of wage schedules in the long run.

7. Threshold Effects

7.1 Introduction

In Section 5 we observed that, under certain circumstances, signals disappeared. This phenomenon leads us to the question of whether changing the parameters of the problem also causes signals to disappear. In this section, we show that not only do signals disappear, but also signaling ceases altogether when the parameters cross threshold points. Three types of parameters will be studied to see if they exhibit threshold effects. We will investigate whether signaling ceases when 1) signaling costs get too high, 2) the variance of the unknown state of the world x gets too small, and 3) signaling noise gets too large.

7.2 Signaling Cost

The first type of parameter to be studied is one affecting the cost of signaling. To illustrate this, we modify the payoff function for the employee population to

$$J_1 = E[\gamma_2 - \alpha c] \quad , \tag{2.13}$$

where $\alpha > 0$ is a scalar cost parameter. For simplicity, the argument will proceed by way of Example 2.3 for $N \geq 3$. With the new payoff J_1 from (2.13), equilibrium conditions (2.10) become

$$x_i = \frac{2\alpha(u_i - u_{i-1})}{x_{i+1} - x_{i-1}} \quad , \qquad i = 1, \ldots, N-1 \quad . \tag{2.14}$$

Since $0 < x_0 < x_{i-1} < x_{i+1} < x_N$,

$$\frac{1}{x_{i+1} - x_{i-1}} > \frac{1}{x_N - x_0}$$

and so

$$x_i > \frac{2\alpha(u_i - u_{i-1})}{x_N - x_0} \quad , \qquad i = 1, \ldots, N-1 \quad . \tag{2.15}$$

As α <u>increases</u>, condition (2.15) will eventually be violated for some i. This implies that a breakpoint has coalesced with another breakpoint or an endpoint (x_0 or x_N). As a particular illustration of this, consider the case of $N = 3$, as shown in Figure 2.10. It can be shown explicitly for this example that as α increases, the breakpoints x_1 and x_2 move away from each other towards the endpoints. More and more people choose the signal u_1. At first it may seem strange that as signaling costs increase, fewer people choose the cheapest signal u_0. To understand this, we must also look at how the wages are changing. First of all, (2.15) can also be written as

$$x_i = \frac{\alpha(u_i - u_{i-1})}{v_i - v_{i-1}} \quad , \tag{2.16}$$

since

$$v_1 - v_0 = \frac{x_2 - x_0}{2}$$

and

$$v_2 - v_1 = \frac{x_3 - x_1}{2} \quad .$$

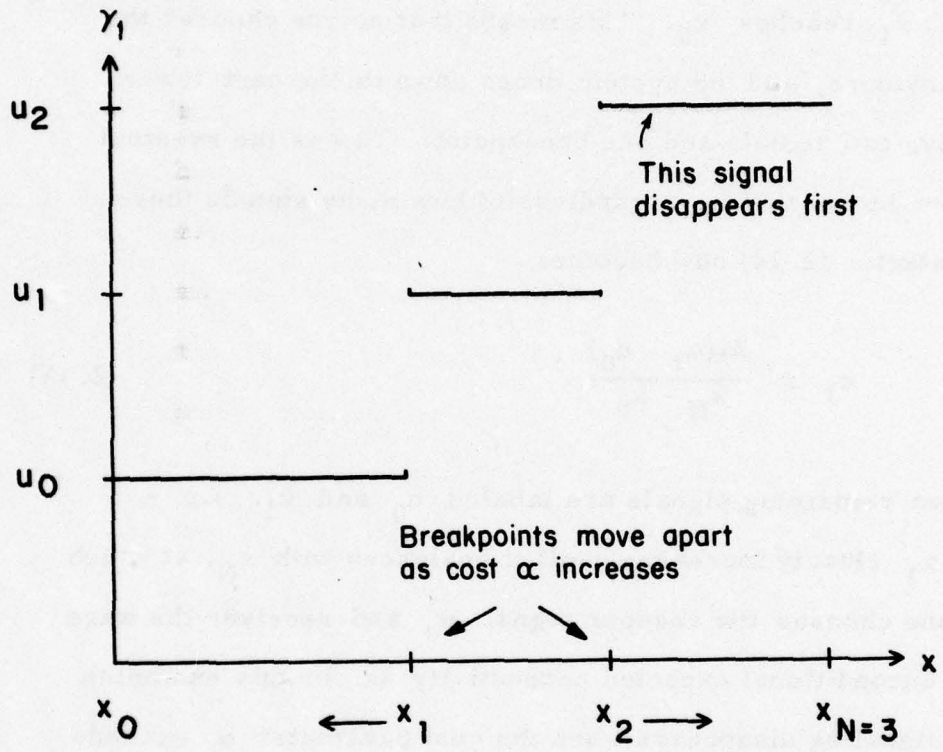


FIG. 2.10 EFFECT OF INCREASING SIGNALING COST

Thus, when x_1 decreases and x_2 increases, both $v_1 - v_0$ and $v_2 - v_1$ also increase. We can deduce from (2.16) that costs α increase faster than $v_2 - v_1$ but slower than $v_1 - v_0$. Therefore, higher ability people switch to u_1 because their costs are rising faster than their relative wages, and lower ability people switch to u_1 for the opposite reason.

Going back to Figure 2.10, it can be shown that x_2 reaches $x_{N=3}$ before x_1 reaches x_0. This means that no one chooses the signal u_2 anymore, and the system drops down to the next lower level, namely, two signals and one breakpoint. This is the eventual outcome from increasing α, regardless of how many signals there were at the start. (2.14) now becomes

$$x_1 = \frac{2\alpha(u_1 - u_0)}{x_N - x_0} \tag{2.17}$$

where the two remaining signals are labeled u_0 and u_1. As α increases, x_1 clearly increases until it coalesces with x_N, at which point everyone chooses the cheaper signal u_0 and receives the wage equal to the unconditional expected productivity ($\bar{x}$ in this example). Therefore, signaling disappears when the cost parameter α exceeds a certain threshold. This result agrees with the intuitive notion that as signaling costs rise, it is no longer worthwhile to invest in the higher educational levels.

Another threshold effect occurs if α is decreased. In this case, the breakpoints will move in the opposite directions and coalesce in the opposite order as before. The $N = 2$ case will again eventually be

reached, but now x_1 will coalesce with x_0, not x_N. Signaling costs get so low that even the lowest ability types choose to pay a little more for u_1 and receive the higher wage v_1. We still have "no signaling," but this solution is inefficient because the employees are overinvesting in the signal. If they all signaled u_0, they could still receive $\gamma_2 = \bar{x}$ but a higher J_1. However, this solution can be considered an equilibrium if we assume that each individual maximizes his own net profit based on the current wages, and that there is no central force (e.g., a union) deciding what is best for the employee population as a whole.

7.3 Variability in the Unknown

Another type of threshold occurs when the variability of the unknown x, the underlying signal, changes. One such parameter is the variance of x, which is proportional to $x_N - x_0$ in the uniform distribution case of Example 2.3. Since $x_N - x_0$ occurs in the denominator of the expressions in (2.15) and (2.17), decreasing $x_N - x_0$ for fixed α has exactly the same effect as increasing α for fixed $x_N - x_0$. Again, the breakpoints shift and coalesce until everyone chooses u_0, and signaling disappears. This result has several intuitive explanations. First of all, it means that as people become more homogeneous, it becomes less important to differentiate them. In other words, the information to be sent through the signal is less worthy of much effort. To see the second meaning, we must observe how the wages are changing. Since

$$v_1 - v_0 = \frac{x_N - x_0}{2}, \tag{2.18}$$

the difference in wages also decreases. The wages would eventually be close enough so that even the individual of highest ability might just as well take the lower (and cheaper) signal u_0, since he cannot receive a significantly higher wage by choosing the higher signal.

However, changing $x_N - x_0$ is more complicated than changing α because not only do the breakpoints move, but so do the endpoints. If x_0 increases, then it may catch up with x_1 before x_1 coalesces with x_N. In this case, everyone would choose the higher signal u_1. To understand the circumstances under which all people choose the more expensive signal, even though the difference in wages is still decreasing, we must analyze three separate cases where $x_N - x_0$ is decreased by shrinking the interval $[x_0, x_N]$.

CASE 1: x_0 fixed and x_N decreased

Since x_0 is fixed, it cannot catch up with x_1, so that x_1 coalesces with x_N.

CASE 2: x_0 increased and x_N decreased at the same rate*

If every time x_0 is decreased by Δ, x_N is increased by Δ, then $\bar{x}$ stays constant and always equals $(x_0 + x_N)/2$. Then from (2.18), $v_1 - v_0 = \bar{x} - x_0$. In order for x_0 to prefer u_0 (and thus maintain two-level signaling), we must have

*Similar analyses can be done if x_0 is increased and x_N decreased at different rates.

$$v_0 - \frac{u_0}{x_0} > v_1 - \frac{u_1}{x_0} \quad ,$$

or

$$u_1 - u_0 > x_0(v_1 - v_0) = x_0(\bar{x} - x_0) \quad . \tag{2.19}$$

A graphical description of the right hand side of (2.19) is shown in Figure 2.11. If $u_1 - u_0 > \bar{x}^2/4$, or if $u_1 - u_0 < \bar{x}^2/4$ and the initial $x_0 > x_0''$,* then (2.19) holds and x_1 eventually coalesces with x_N as x_0 increases. If, on the other hand, $u_1 - u_0 < \bar{x}^2/4$ and $x_0 < x_0'$, then the right hand side of (2.19) increases until (2.19) is violated and x_1 coalesces with x_0.** The intuitive reasons for this are twofold. First of all, $u_1 - u_0$ must be small enough so that there is not so much difference in cost between the signals. Secondly, x_0 must be sufficiently small, so that the average productivity (also wage) for the lower group, namely v_0, is then much smaller than v_1, so that by the time v_0 comes close to v_1, the lowest ability group has already decided that v_1 is enough of an inducement to choose u_1.

CASE 3: x_0 increased and x_N fixed

From (2.18), two-level signaling continues if

$$2(u_1 - u_0) > x_0(x_N - x_0) \quad . \tag{2.20}$$

---

*As shown in Figure 2.11, x_0' and x_0'' are defined as

$$x_0''(\bar{x} - x_0'') = x_0'(\bar{x} - x_0') = u_1 - u_0, \quad x_0' < x_0'' \, .$$

**If $x_0 < x_0'$, then initially $x_1 = (u_1 - u_0)/(\bar{x} - x_0) < x_0' = (u_1 - u_0)/(\bar{x} - x')$. When x_0 catches up with x_1 at the value x_0', x_1 cannot have coalesced already with x_N, because $x_0' < \bar{x} < x_N$.

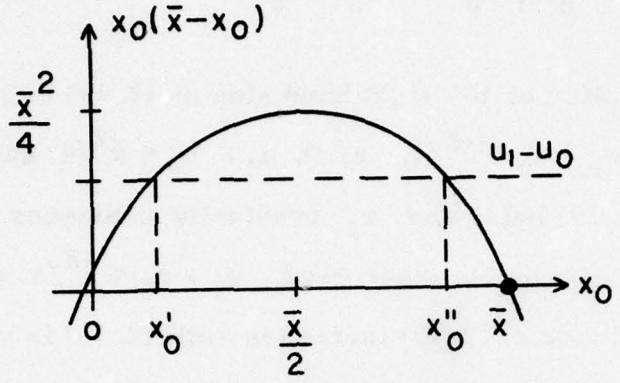


FIG. 2.11 CASE 2 FOR CHANGING $[x_0, x_N]$

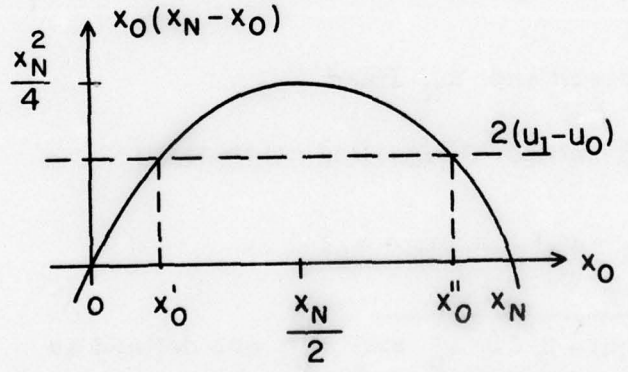


FIG. 2.12 CASE 3 FOR CHANGING $[x_0, x_N]$

A graphical description of the right hand side of (2.20) is shown in Figure 2.12. By an analysis similar to that of Case 2, x_1 coalesces with x_N unless $2(u_1-u_0) < x_N^2/4$ and $x_0 < x_0'$. The same intuitive arguments hold as in Case 2.

The conclusion from the preceding discussion is that decreasing the range $[x_0, x_N]$ by increasing x_0, decreasing x_N, or both results in "no signaling." The parameters of the problem determine whether the employees *chose the higher or lower signal.*

A similar analysis in Appendix II-A describes what happens when $x_N - x_0$ increases. In general, x_1 will decrease and coalesce with x_0, resulting in "no signaling." However, depending on other parameters, x_0 may decrease faster than x_1, so that x_1 never catches up to x_0. Differentiated signaling continues until x_0 reaches zero (recall that x_0 was assumed to be positive).

7.4 Signaling Noise

The third type of parameter we want to investigate is signaling noise. Suppose now that the employer has a noisy measurement of education and observes $y = u + \epsilon$ instead of u, where ϵ is the noise. Then his strategy is a function of a noise-corrupted signal:

$$v = \gamma_2(y) = \gamma_2(u + \epsilon) = \gamma_2(\gamma_1(x) + \epsilon) \quad .$$

The equilibrium condition remains the same as before, that is

$$\gamma_2(y) = E_{/y}(s) \quad . \tag{2.21}$$

It is difficult to give an economic interpretation to ϵ. If we assume that "educational level" reflects a ranking of a composite measure of years of education, performance, courses, and quality of the school, and we assume that this ranking is known to employer and employees alike, then it appears that "noise" can only mean interference in the communication link between what the individual does and what the employer observes. But if the job application form is complete enough and the individuals do not lie, then noise, in this sense, should be eliminated. However, we can still treat ϵ as a purely mathematical entity. (In the next chapter, noise will play a more important role.)

Continuing to use Example 2.3 to illustrate the main ideas, we assume ϵ has a uniform distribution between some -b and b. For the case of N = 2, the employees' strategy γ_1 remains a step function with two signaling levels, as in the case of no signaling noise. However, as will be shown next, the employer's strategy γ_2 does <u>not</u> remain a two-step wage schedule. To see this, refer to Figure 2.13, where γ_2 is plotted vs. $\underline{y}$, not u. Assuming u_0 and u_1 are fixed signals, any y in an interval of $\pm b$ around u_0 and u_1 could be observed by the employer. If a y between $u_0 - b$ and $u_1 - b$ is observed, the employer knows that u_0 was signaled, so that x must be between x_0 and x_1. The wage v_0 is the average productivity for that interval, namely, $(x_0 + x_1)/2$ for this example. Similar arguments can be made to determine v_1 and v_2, as shown in Figure 2.12. Therefore, the two-step wage schedule becomes a three-step schedule if the noise is

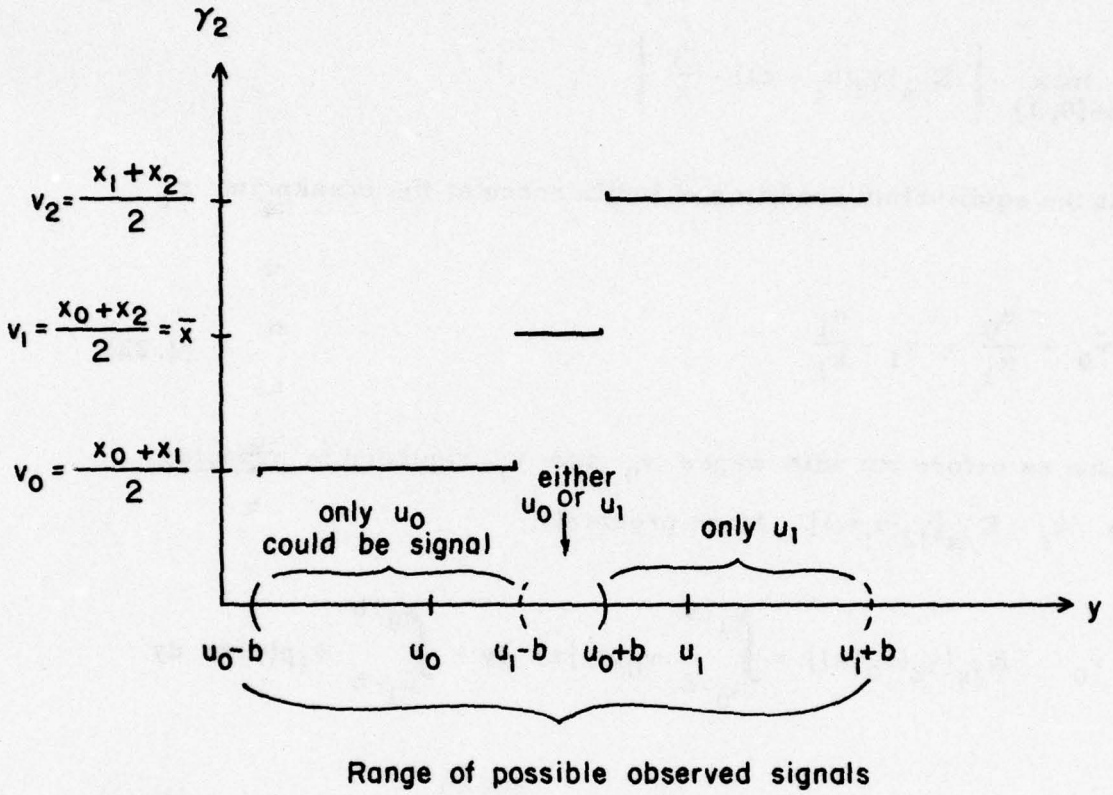


FIG. 2.13 WAGE SCHEDULE WITH SIGNALING NOISE

uniformly distributed and there is a central overlap region of uncertainty.*

Given the three-step wage schedule, the individuals now want to maximize expected net profit, where the expectation is taken over both x and y. This means

$$\max_{i\in\{0,1\}} \left\{ E_{/x}[\gamma_2(u_i + \epsilon)] - \frac{u_i}{x} \right\} ,$$

so that the equilibrium condition of indifference at the breakpoint x_1 is

$$\bar{v}_0 - \frac{u_0}{x_1} = \bar{v}_1 - \frac{u_1}{x_1} , \qquad (2.22)$$

the same as before but with wages v_0 and v_1 replaced by <u>expected</u> wages $\bar{v}_i = E_{/x}[\gamma_2(u_i+\epsilon)]$. More precisely,

$$\bar{v}_0 = E_{/x}[\gamma_2(u_0+\epsilon)] = \int_{u_0-b}^{u_1-b} v_0 p(y|x)\ dy + \int_{u_1-b}^{u_0+b} v_1 p(y|x)\ dy$$

$$= v_0 \frac{\Delta u}{2b} + v_1 \frac{2b-\Delta u}{2b} , \qquad (2.23)$$

where $\Delta u = u_1 - u_0$. Similarly,

$$\bar{v}_1 = E_{/x}[\gamma_2(u_1+\epsilon)] = v_1 \frac{2b-\Delta u}{2b} + v_2 \frac{\Delta u}{2b} . \qquad (2.24)$$

---

*For the case of Gaussian noise, the wage schedule is a continuous function, not a step function. However, the breakpoint equilibrium conditions can still be determined. The details are complicated, and so are omitted here.

Thus, the expected wage $\bar{v}_i$ is just a weighted average of wages v_i and v_{i+1}. From (2.23) and (2.24), (2.22) becomes

$$\frac{\Delta u}{2b} v_0 - \frac{u_0}{x_1} = \frac{\Delta u}{2b} v_2 - \frac{u_1}{x_1}$$

$$\Rightarrow \quad x_1 = \frac{1}{\frac{1}{2b}(v_2 - v_0)} = \frac{4b}{x_2 - x_0} \quad . \tag{2.25}$$

(Also, we must have $x_0 < x_1 < x_2$.)

EXAMPLE 2.5. Let

$$b = .6, \qquad x_0 = 1, \qquad x_2 = 2.5 \quad .$$

Then

$$v_0 = 1.33$$

$$v_1 = 1.75$$

$$v_2 = 1.92$$

and

$$x_1 = 1.6 \quad .$$

It is easy to see from (2.25) that as the signal becomes noisier and b increases, x_1 increases and, in general, coalesces with x_2. Everyone chooses the cheaper signal u_0. Therefore, as we would expect, if the signal is too noisy, signaling will cease. However, x_1 may not coalesce with x_2. Referring to Figure 2.13, we see that as b increases, the interval of y's which are paid $v_1 = \bar{x}$ expands. The

other intervals remain constant in size, but the left interval shifts to the left and the right interval to the right. Eventually, negative values of y will appear. Since negative signals do not make sense from an economic point of view, we stop increasing b when $u_0 - b = 0$. If b increases to u_0 before x_1 reaches x_2, then $u_0 - b = 0$, and differentiated signaling remains in effect.

If b <u>decreases</u>, x_1 decreases and coalesces with x_0. This result is surprising, since it says that when the noise is small enough, everyone chooses the more expensive signal. To understand this, we must again look at how the difference in the wages is changing:

$$\bar{v}_1 - \bar{v}_0 = \frac{\Delta u}{2b} (v_2 - v_0) \quad .$$

Thus, as b decreases, $\bar{v}_1 - \bar{v}_0$ increases. Eventually, the expected wage $\bar{v}_1$ will be large enough, so that choosing the higher signal becomes worthwhile.

7.5 Summary of Threshold Effects

The results of this section will now be summarized.* First of all, if the signaling cost parameter α is increased, then the cost of education gets too high, and everyone chooses the cheapest signal. If α is decreased, then the opposite happens: cost of education becomes so low that it becomes worthwhile to pay a little more for the higher signal and receive the higher wage.

---

*All of the statements are in reference to Example 2.3.

Changing the variability of the unknown state of the world x is more complicated than changing α. The parameter in the former case is the length $x_N - x_0$ of the interval $[x_0, x_N]$. The results are summarized in Table 2.3. In general, if the variability is too small, then the individuals are not differentiated enough to make it worthwhile for them to signal, so that everyone chooses the cheaper signal u_0. However, if the difference in the wages decreases slower than the difference in the costs for the lowest ability group x_0, then the higher wage is enough of an inducement for all individuals, even the lowest ability types, to choose the more expensive signal. If the variability increases,* we would expect people to continue signaling, since they are becoming more dissimilar. However, again the results depend on how the difference in the wages is changing. If it is increasing faster than the difference in costs for the lowest ability group, then eventually everyone will choose the higher signal in order to receive the higher wage. Therefore, the threshold effects for the variability of x parameter depend on other parameters of the problem.

The last parameter is signaling noise. It was shown that if the noise is too high, then it becomes too difficult for the employer to determine the ability from the signal, so that everyone chooses the cheaper educational level. If the noise is small, then the signal is a better indication of ability, but a secondary effect takes over. The expected wage from the more expensive signal is sufficiently high to induce everyone to choose that signal.

---

*The details for increasing $x_N - x_0$ are discussed in Appendix II-A.

TABLE 2.3

Threshold Effects from Changing $x_N - x_0$

	Decreasing $x_N - x_0$	Increasing $x_N - x_0$ (see Appendix II-A for details)
Case 1: x_0 fixed	$x_1 \to x_N$ All choose u_0.	$x_1 \to x_0$ All choose u_1
Case 2: same rate	$x_1 \to x_N$, unless $u_1 - u_0$ and x_0 are sufficiently small so that $v_1 - v_0$ decreases slower than $\frac{u_1 - x_0}{x_0}$ in such a way that x_0 catches up with x_1 before x_1 reaches x_N. Then $x_1 \to x_0$.	Differentiated signaling continues as $x_0 \to 0$, unless $u_1 - u_0$ sufficiently small and x_0 sufficiently large so that $v_1 - v_0$ increases faster than $\frac{u_1 - u_0}{x_0}$ in such a way that x_1 catches up with x_0. Then $x_1 \to x_0$.
Case 3: x_N fixed	Same argument as in Case 2.	Same argument as in Case 2.

REFERENCES:

1. M. Spence, "Competitive and Optimal Responses to Signals: An Analysis of Efficiency and Distribution," J. Econ. Theory, Vol. 7, pp. 296-332.

2. M. Spence, Market Signaling: Informational Transfer in Hiring and Related Screening Processes. Cambridge, MA, Harvard University Press, 1974.

3. M. Spence, "Competition in Salaries, Credentials, and Signaling Prerequisites for Jobs," Q. J. Econ., Vol. XC, pp. 51-74, 1976.

4. J. Riley, "Competitive Signaling," J. Econ. Theory, Vol. 10, pp. 174-186, 1975.

5. Y. C. Ho and K. C. Chu, "Information Structure in Many-Person Optimization Problems," Automatica, Vol. 10, pp. 341-351, 1974.

6. A. V. Malishevskii, "Models of Joint Operation of Many Goal-Oriented Elements: Part I," Automation and Remote Control, Vol. 11, pp. 92-110, 1972.

APPENDIX II-A

THRESHOLD EFFECTS FOR INCREASING VARIABILITY OF x

Increasing $[x_0, x_N]$ for $c = u/x$, $p(x)$ uniform

The arguments in this section are completely analogous to those in Section 7.3 and will refer to equations and figures from there. (See Table 2.3 for a summary.) It is easy to see from (2.17) that if $x_N - x_0$ is increased, then x_1 decreases and, in general, coalesces with x_0. Everyone chooses the more expensive signal u_1. This is because, from (2.18), $v_1 - v_0$ increases. As v_1 becomes significantly higher than v_0, it eventually becomes worthwhile for even the lowest ability group to choose the higher signal. However, this result again depends on $u_1 - u_0$. It may be that v_1 can never be sufficiently large to attract all employees. This means that even though x_1 decreases as $x_N - x_0$ increases, x_0 decreases faster than x_1. x_1 can never catch up to x_0, so that differentiated signaling continues.* Again, we must consider three cases:

CASE 1A: x_0 fixed and x_N increased

Since x_0 is fixed, x_1 can catch up to it.

CASE 2A: x_0 decreased and x_N increased at the same rate

Referring to Figure 2.11, in a manner completely analogous to Case 2 in Section 7.3, we see that if $u_1 - u_0 > \bar{x}^2/4$, or if $u_1 - u_0 < \bar{x}^2/4$

*Since $x_0 > 0$, the system must be stopped before x_0 reaches 0.

but the initial $x_0 < x_0'$, then the breakpoint x_1 decreases but never catches up with x_0. The employees continue to signal at different levels. If $u_1 - u_0 < \bar{x}^2/4$ and $x_0 > x_0''$, then as x_0 decreases, the right hand side of (2.19) increases until (2.19) is violated, x_1 coalesces with x_0, and signaling disappears. Intuitively, if $u_1 - u_0$ is sufficiently small, then as $v_1 - v_0$ increases, it will eventually be worthwhile for the lowest ability group to pay a little more for u_1 and receive the much higher wage v_1. If the initial x_0 is sufficiently large, then $c = u/x_0$ will not be too large, so that it will again be worthwhile to purchase u_1.

CASE 3A: x_0 decreased and x_N fixed

Figure 2.12 describes the situation here. Signaling continues unless $2(u_1 - u_0) < x_N^2/4$ and $x_0 > x_0''$, in which case x_1 coalesces with x_0. The same intuitive arguments hold as in Case 2A.

Thus, in the case of increasing the range $[x_0, x_N]$, signaling may or may not disappear, depending on other parameters in the problem.

CHAPTER III

SIGNALING AND INFORMATION THEORY

1. Introduction

In the previous chapter, an example of a signaling problem was analyzed in the context of economic theory. This chapter will analyze signaling in the context of Shannon theory, also sometimes referred to as classical information theory. This theory forms the foundation for the following standard communication problem: send a message, or signal, through a noisy channel so as to minimize the amount the signal is distorted. We will show how the main components of this problem can be captured in a team theory formulation (to be defined below) with a signaling information structure. Figure 3.1 summarizes the connection between this problem and the Spence problem.

Before going on to the next section, we need to describe what a team problem is and how it relates to the Spence problem. A decision and control problem is called a "team" problem when there is more than one decision maker, each DM has different information, but all DMs have the same objective function J. The Spence problem was not a team problem, since $J_1 \neq J_2$, but was an example of a "nonzero-sum (NZS) game" (called "nonzero-sum" because $J_1 + J_2 \neq 0$). An "optimal" solution in the Spence problem was characterized as a Nash equilibrium, as defined in Section II.2. On the other hand, a "team optimal" strategy pair (γ_1^*, γ_2^*) is defined as (where J is to be minimized):

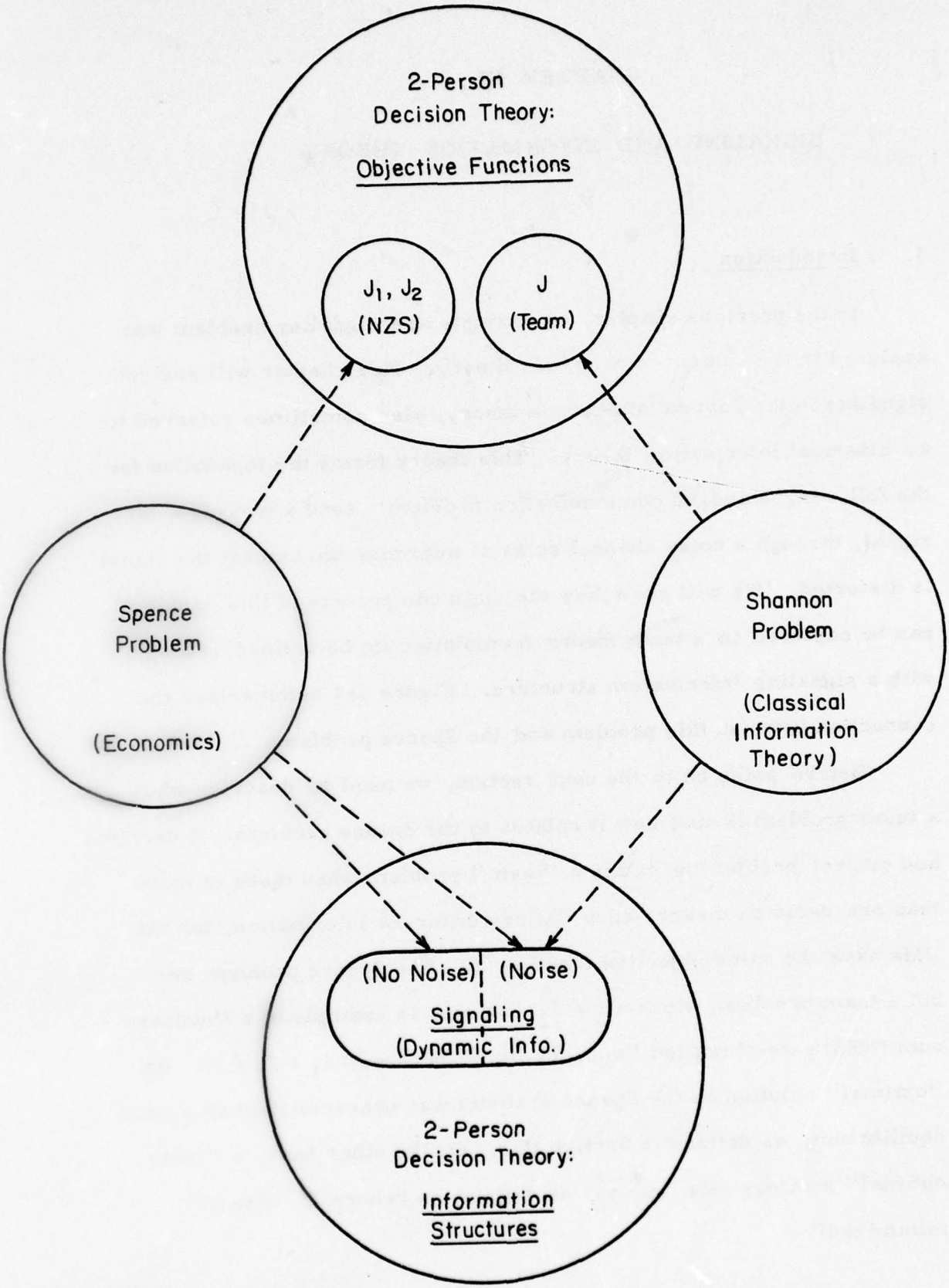


FIG. 3.1 COMPARISON OF SPENCE AND SHANNON PROBLEMS

$$J(\gamma_1^*, \gamma_2^*) \leq J(\gamma_1, \gamma_2) \quad \forall \text{ admissible } (\gamma_1, \gamma_2) \quad .$$

As shown by Radner [7], the first order necessary and sufficient conditions are the same as for a Nash equilibrium, but the second order convexity conditions are not. However, the second order conditions in [7] are only for the case of static information structure, and cannot be carried over to dynamic information. This is because γ_2 is a function of γ_1, so that the convexity of $J(\gamma_1, \gamma_2(\gamma_1))$ in both γ_1 and γ_2 cannot be determined until γ_2 is specified. Therefore, "convexity" of an objective function when the information is dynamic is yet to be defined.

2. Communication System as Team Problem

The major problem in communication theory is to send information from a source through a channel to a receiver in the most "reliable" way, where "reliability" is yet to be defined. Wyner [14] says that, in general, there are two limitations on the reliability of the communication system. First of all, the channel may have noise, such as static in a radio channel. The second limitation is what Wyner calls "source-channel mismatch." For example, the source may emit binary symbols, as with a computer, whereas the channel may be able to accept only continuous data, as with a radio. Also, the rate at which the source emits data may be different from the rate at which the channel can process it. To combat these limitations, an encoder is placed between the source and channel, and a decoder between the

channel and receiver. Figure 3.2 illustrates these ideas for the standard model of a communication system* (see [8], [14]). First of all, a (memoryless**) source emits source symbols, or messages, denoted as x, at a rate ρ_s symbols/second, which go into the encoder and come out as signals, or codewords, u. The functional relationship between x and u as defined by the encoder is denoted as $u = \gamma_1(x)$. Next, the signals u are transmitted through a (memoryless) channel that processes inputs at a rate ρ_c symbols/second, and may be corrupted by noise ϵ. We will be concerned with additive noise, so that the received signal y coming out of the channel will be expressed as $y = u + \epsilon$. Lastly, the decoder converts y to symbols v, and sends them to the receiver. Since the goal of the decoder is to determine which message x was originally sent, we will also write v as $\hat{x}$, so that the decoder is, in some sense, the inverse of the encoder. As with the encoder, the decoder is defined by some function γ_2, so that $v = \gamma_2(y) = \gamma_2(u + \epsilon) = \gamma_2(\gamma_1(x) + \epsilon)$. We immediately see how similar this looks to the Spence problem, where γ_2 is a noise corrupted function of γ_1. More will be said about this below.

To complete the description of the channel, $p(y|u)$, the transitional probability density of y given u, and a cost function $\phi(u)$ must be specified; for example, $\phi(u) = u^2$. To complete the description of the source, $p(x)$, the probability density for the source output, and

---

*In order to relate this problem to the Spence problem later on, we will use notation that matches the notation in the previous chapter, not the notation that necessarily occurs in the information theory literature.

**The memoryless assumption means, in general, that current behavior does not depend on the past.

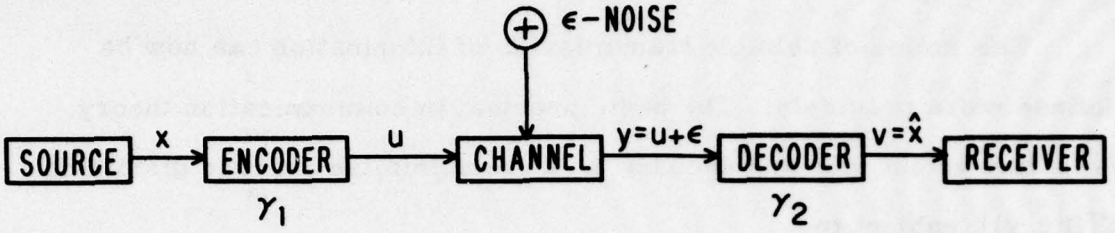


FIG. 3.2 COMMUNICATION SYSTEM

$D(x, v)$, the distortion function, must be specified; for example, $D(x, v) = (x - v)^2$. Distortion is a measure of how v differs from x, regardless of whether the signal has been sent through a channel or not. An example of where nonzero distortion occurs without the presence of a channel is in data compression, where less significant information is deleted or condensed in order to transmit more significant information more reliably.

The notion of reliable transmission of information can now be defined more precisely. The basic problem in communication theory is to find an encoder and decoder so as to minimize average distortion $E[D(x, v)]$ subject to

$$E[\phi(u)] \leq \alpha \quad , \tag{3.1}$$

where α is some fixed constant. Inequality (3.1) is a constraint on the amount of signal power, where by "signal" we mean the channel input variable u. Since minimizing distortion is the single goal, the problem lends itself naturally to the following team formulation:

$$\min_{\gamma_1, \gamma_2} J = \underset{\substack{\uparrow \\ \text{distortion}}}{E[D(x, v)]} = \underset{\substack{\uparrow \\ \text{signaling}}}{E[D(x, \gamma_2(\gamma_1(x)+\epsilon))]} \text{ s.t. } \underset{\substack{\uparrow \\ \text{power constraint}}}{E[\phi(\gamma_1(x))]} \leq \alpha \, , \tag{3.2}$$

that is, minimize average distortion subject to a power constraint as the encoder and decoder are varied. Thus, DM1 is the encoder with strategy $\gamma_1(x) = u$, and DM2 is the decoder with strategy $\gamma_2(y) = v$. Since $v = \gamma_2(u + \epsilon) = \gamma_2(\gamma_1(x) + \epsilon)$, this problem exhibits precisely

what we set out to show, namely, signaling with noise, since γ_2 is a function of a noise-corrupted γ_1. This particular notation shows how similar this problem is to the Spence problem. In fact, if $D(x, v) = (x - v)^2$, then $J = E[D(x, v)]$ is precisely J_2 from the Spence problem.

Wyner [14] posed the communication problem in language which easily transfers to the team problem in (3.2), although he does not explicitly mention team theory. Witsenhausen [12], however, recognized that the communication problem could be formulated as a team problem* with a dynamic information structure, but did not investigate this further. Whittle and Rudge [10] took the opposite point of view. They started with a team problem and showed that it could be interpreted as a communication problem. Their team problem was a more general version of (3.2), where x, u, ϵ, etc. represented infinite time sequences,** so that they could use the results of information theory to solve for the optimal value of (3.2).

Now that the communication problem has been reduced to a team problem, several questions from a decision and control point of view arise. First of all, an obvious question is: what is the team optimal strategy pair (γ_1^*, γ_2^*) for the problem in (3.2)? Once this pair has been determined, a second obvious question is: what is the value of the optimal objective $J^* = J(\gamma_1^*, \gamma_2^*)$? Our immediate response might be to

---

*He called it a "nonclassical stochastic control problem."

**It will be shown later why the assumption of infinite sequences is important in information theory.

despair of ever answering these questions because of all the difficulties associated with team problems with dynamic information, as discussed in Section 1. In fact, Witsenhausen [11] extensively studied a similar team problem without answering these questions. Fortunately, since our team problem was motivated by a communication system, we can take the same approach as in Whittle and Rudge, and use the results of Shannon information theory to answer some, but not all, of these questions.

If γ_1^* and γ_2^* cannot be found, or if they are very complicated functions, then a next question to ask might be whether there are suboptimal strategies whose objective J does not differ too much from J^*, but which are easier to compute than γ_1^* and γ_2^*. For example, when are linear strategies, which are simple to express, optimal, and when are they not optimal?

In the team formulation of (3.2), γ_1 and γ_2 could be mappings between <u>scalar</u> variables. However, if the admissible strategy spaces for γ_1 and γ_2 were expanded to include mappings between <u>vectors</u>, then increasing the dimensions of x, u, ϵ, etc. might result in lower distortion than if the variables were restricted to being just scalars. Certainly, by increasing the strategy space, we cannot do worse and may, in fact, do better. *Therefore, this observation leads us to ask* how the dimensions of the variables x, u, etc. affect the solutions (γ_1^*, γ_2^*) and J^*. As mentioned above, in information theory, these variables represent infinite sequences, so that they can be thought of as infinite-dimensional vectors. Thus, information theory might be able to tell us something about the affect of dimensionality on the solution.

To summarize, the questions we want to answer are:

QUESTIONS:

1. What are the team optimal strategies (γ_1^*, γ_2^*)?
2. What is the value of the optimal objective $J^* = J(\gamma_1^*, \gamma_2^*)$?
3. Are there suboptimal strategies that are easy to implement, and how does their objective J differ from J^*?
4. How do the dimensions of the variables x, u, etc. affect the solutions (γ_1^*, γ_2^*) and J^*?

Before we address them, we momentarily digress from our team theory point of view to define the basic concepts and results from Shannon information theory. (For more detail, see [1], [2], [8], [14], and [6].)

3. Shannon Theory*

3.1 Basic Concepts

Shannon theory provides the theoretical foundation for communication theory by establishing an upper bound, called "channel capacity" (C), on the amount of information that can be transmitted through a channel. It also provides a quantitative measure of information that can be used in characterizing not only the channel capacity, but also the rate at which the source produces information, called the "source rate (R)."

*Those familiar with information theory may wish to skip this section, since its purpose is solely to educate people, such as economists and control theorists, who have little or no knowledge of information theory. The intent is not to shed new light on Shannon's results, but rather to define terminology and concepts for later use.

Intuitively, if R, the rate at which the source produces information, is less than C, the maximum rate the channel can process information, we would expect that the source and channel could be joined in some way to produce a communication system that transmits information at a rate R. This is exactly what Shannon's Coding Theorem says, namely, that if $R \leq C$, then there exists an encoder and decoder joining the source-receiver pair to the channel such that information can be transmitted at a rate as close to C as desired with arbitrarily small probability of error, in the limit as the length (or duration) of the encoded messages gets sufficiently large. If $R > C$, then the source is producing information faster than the channel can process it, so that a certain amount of error is unavoidable. These intuitive ideas will be made more precise later on when we return to the Coding Theorem in more detail. Before we define what is meant by "rates of information," we must define the concept of "information" first. Shannon's abstract measure of information, to be described next, is interesting, but, by itself, does not provide any new results. Its real importance lies in the fact that, with this measure, the important Coding Theorem could be proved.

The randomness inherent in the messages and signals of a communication system implies that information is statistical, so that any measure of it must involve probabilities. As mentioned in the previous section, the particular probabilities* required are the source

*The discussion and definitions to follow will all be for the case of a discrete source and channel, so that probabilities instead of densities will be used. The information measure for the continuous case can be similarly defined, but is more complicated to interpret and so will be omitted here.

output and channel transition probabilities. Therefore, these probabilities will play a role in the definitions to follow.

Since information can be defined in a purely statistical sense, the definitions will first be stated in terms of abstract sets of events and then interpreted in terms of a source and channel. To simplify matters, think of a source generating symbols a_i from a discrete, finite set A, which are then input directly into a channel, emerging as output symbols b_j from a discrete, finite set B. In an abstract sense, A and B are just random variables characterized by probabilities $\{p(a_i)\}$ and $\{p(b_j)\}$, respectively. Then we have the following definitions:

1. Information: $I(a_i) = \log 1/p(a_i) = -\log p(a_i)$
   = amount of information received if told event a_i has occurred.

Intuitively, if $p(a_i)$ is small, then a lot of information is received if the unlikely event a_i has occurred. If the log is in base 2, then the unit of $I(a_i)$ is called a "bit." If it is in base e, then the unit is a "nat." We will be using the "bit" notation in the rest of this chapter. The particular choice of "log" comes about because it satisfies certain desirable axioms. See [8] for a detailed discussion of these axioms, and [15] for alternatives to log as the information measure.

2. Entropy (bits/symbol): $H(A) = E[I(A)] = \Sigma_i\ p(a_i)I(a_i)$
   = average amount of information received after being told what the source emitted
   = average prior uncertainty regarding what the source will emit

= average number of "yes-no" questions to be answered to determine output

= rate at which source produces information subject to no distortion (to be explained later).

3. Conditional Entropy: $H(A|b_j) = E_{a/b_j}[I(A|b_j)] = \Sigma_i p(a_i|b_j) I(a_i|b_j)$

= average information from A given observation b_j .

4. Equivocation: $H(A|B) = E_b E_{a/b}[I(A|b)] = \Sigma_j p(b_j) H(A|b_j)$

= average information from A given output is observed

= average uncertainty of what source emitted after observing an output signal

= average amount of information missing in the received signal

= average amount of additional information that must be supplied per second at the receiving point to correct the received message.

5. Mutual Information: $I(A, B) = H(A) - H(A|B)$

$$= \sum_i p(a_i) \log \frac{1}{p(a_i)} - \sum_j p(b_j) \sum_i p(a_i|b_j) \log \frac{1}{p(a_i|b_j)}$$

$$= \sum_j \sum_i p(a_i, b_j) \log \frac{p(a_i|b_j)\, p(b_j)}{p(a_i)\, p(b_j)}$$

$$= \sum_i \sum_j p(a_i, b_j) \log p(a_i, b_j)/p(a_i)p(b_j)$$

= average information provided by observing one output

= rate of transmission of information through the channel

= measure of statistical dependence between A and B (the more dependent they are, the more information we get about A from observing B).

The important consequence of these concepts is that now the source rate R and channel capacity C can be expressed as solutions to a pair of optimization problems involving mutual information. For a general channel with power constraint (3.1), capacity can be written as (using the notation from Figure 3.2)

$$C \triangleq C(\alpha) = \rho_c \sup_{p(u)} I(u,y) \quad \text{s.t.} \quad E[\phi(u)] \leq \alpha \ , \tag{3.3}$$

where the supremum is taken over all input probabilities satisfying the constraint, and $C(\alpha)$ is in units of bits/second. As mentioned earlier, $C(\alpha)$ is defined as the maximum rate that information (in bits) can be sent through the channel <u>essentially error-free</u> ("essentially" in the sense that the probability of error can be made arbitrarily small).

Similarly, for a general source, the rate can be expressed as

$$R \triangleq R(\beta) = \rho_s \inf_{p(v/x)} I(x,v) \quad \text{s.t.} \quad E[D(x,v)] \leq \beta \ , \tag{3.4}$$

where

$$E[D(x,v)] \leq \beta$$

is called the <u>fidelity criterion</u> and $R(\beta)$ is called the <u>rate distortion</u>

function, with β a nonnegative constant. At first it might seem strange to talk about minimizing a rate, since we always talk about maximizing the transmission rate. But as Berger [2] points out, with rate distortion functions, it is the source-receiver pair that is given, not the channel. What is being minimized is, in some sense, the time and effort it takes to code a message. Thus, as proved by Shannon, $R(\beta)$ can be interpreted as the minimum number of binary digits per second required to represent a message, subject to distortion no more than β. If the source-receiver pair is to be linked to a channel, then $R(\beta)$ can also be interpreted as the minimum capacity that channel must have. $R(\beta)$ is a decreasing function of β, since a higher distortion allowed means fewer binary digits needed to represent the message. This is easy to see mathematically, since larger β means expanding the set of admissible $p(v|w)$ over which the infimum is taken. Since entropy is the rate at which information is generated subject to no distortion, the rate distortion function is just a generalization of the concept of entropy. For discrete sources, $R(0) = \rho_s H(w)$. For continuous sources, such as Gaussian, $R(0) = \infty$, since a real number would require an infinite number of bits to represent it perfectly.

The following are some examples of $C(\alpha)$ and $R(\beta)$ from Wyner [14], derived directly from the definitions (3.3) and (3.4), respectively.

EXAMPLE 3.1: Suppose we have a binary source such that $\Pr\{X = 0\} = \Pr\{X = 1\} = 1/2$, and the distortion function $D(x, v) = 0$

if $x = v$ and $D(x, v) = 1$ if $x \neq v$ (this implies that $E[D(x, v)] = P_e$, where P_e is the probability of error). Then

$$R(\beta) = \begin{cases} \rho_s(1-h(\beta)), & 0 \leq \beta \leq \frac{1}{2} \\ 0 \quad , & \beta \geq \frac{1}{2} \end{cases} \tag{3.5}$$

where $h(\beta) = -\beta \log_2 \beta - (1 - \beta) \log_2 (1 - \beta)$, $0 < \beta \leq 1/2$ and $h(0) = \lim_{\beta \to 0} h(\beta) \triangleq 0$. From the definition of entropy and $h(\beta)$, we see that $h(1/2) = H(x) = 1$, and $R(0) = \rho_s$ (see Figure 3.3). The reason $R(\beta) = 0$ for $\beta \geq 1/2$ is because distortion $\beta = 1/2$ can be attained by always guessing $v = \hat{x} = 0$. That is, the decoder output is a stream of zeros, regardless of the input, so that no information is being produced.

EXAMPLE 3.2: Consider now a Gaussian source where x has a Gaussian density function with zero mean and variance σ^2, and the distortion function is $D(x, v) = (x - v)^2$. Then

$$R(\beta) = \begin{cases} \frac{\rho_s}{2} \log \frac{\sigma^2}{\beta} , & 0 \leq \beta \leq \sigma^2 \\ 0 \quad , & \beta \geq \sigma^2 \end{cases} \tag{3.6}$$

(See Figure 3.4.) The reason $R(\beta) = 0$ for $\beta \geq \sigma^2$ is because $\beta = \sigma^2 = \text{variance } (x)$ can be attained by guessing that x is the prior mean; i.e., $v = \hat{x} = 0$ for this example. Again, this means that the decoder output is all zeros, and no information is being produced.

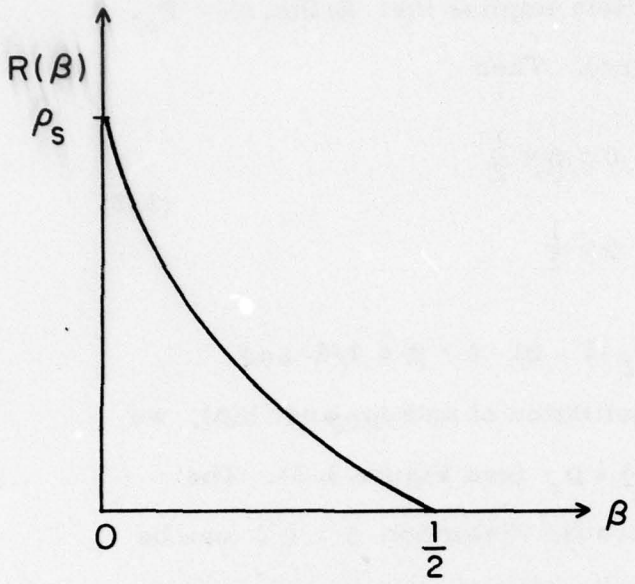


FIG. 3.3 R(β) FOR BINARY SOURCE

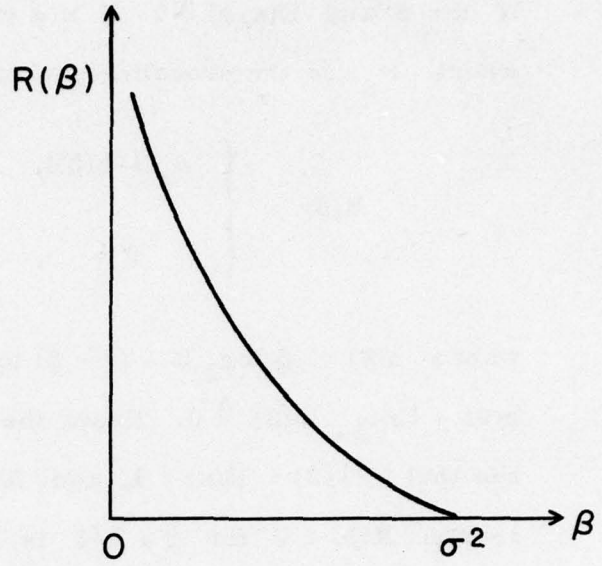
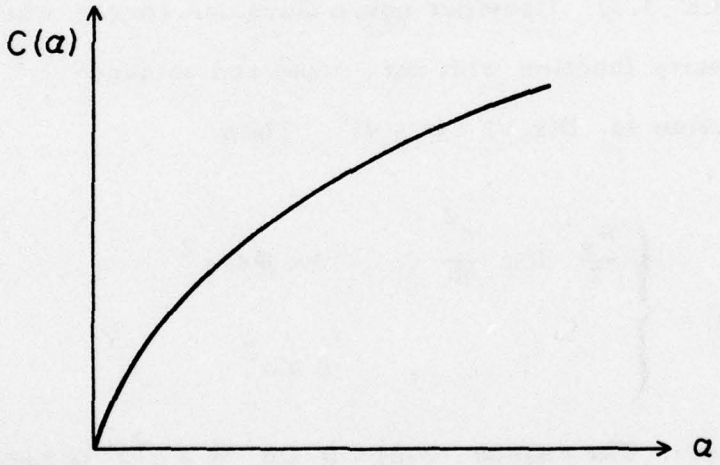


FIG. 3.4 R(β) FOR GAUSSIAN SOURCE



FIG, 3.5 C(α) FOR GAUSSIAN CHANNEL

EXAMPLE 3.3: Consider a Gaussian channel where noise ϵ has a Gaussian density with zero mean and variance σ^2, and the cost function is $\phi(u) = u^2$. Then

$$C(\alpha) = \frac{\rho_c}{2} \log \left(1 + \frac{\alpha}{\sigma^2}\right) , \tag{3.7}$$

where α/σ^2 can be considered the signal-to-noise ratio. Figure 3.5 shows that C is an increasing function of α; that is, capacity can be increased if more channel input power is allowed.

Although $R(\beta)$ and $C(\alpha)$ were defined in terms of single input and output symbols, actual coding does not, except in rare circumstances, involve immediately sending each source symbol through the channel, even if $\rho_s = \rho_c$. In order to combat noise limitations and source-channel mismatch (such as when $\rho_s \neq \rho_c$), the encoder waits for many source symbols and then codes them altogether. A wider range of codes is then available to the encoder, so that cleverer codes can be constructed. Similarly, the decoder waits for many channel outputs before it decodes. For example, if $\rho_s \neq \rho_c$, then the source and channel are *not* synchronously compatible. In order to match them, let the encoder wait T seconds until $n = \rho_s T$ symbols have been emitted. In this time, the channel can process $N = \rho_c T$ symbols. Thus, let x be an n-vector and u an N-vector; this is called <u>block coding.</u>* The new vector source is called the <u>n-th extension</u> of the original source, and the <u>N-th extension</u> of the channel can be similarly

*Block coding is used not only for synchronization, but also to combat noise.

defined. The "new" source symbol rate is now ρ_s/n (in units of n-vectors per second), which equals the "new" channel rate ρ_c/N (N-vectors/second). Assuming that the vector components are independent, we define

$$p(x) = \prod_{i=1}^{n} p(x_i)$$

$$p(y|u) = \prod_{i=1}^{N} p(y_i|u_i) \quad .$$

Also, D_n, the distortion of the n-th extension of the source, is

$$D_n = \frac{1}{n} \sum_{i=1}^{n} D(x_i, v_i) \quad , \tag{3.8}$$

and ϕ_N, the cost of the N-th extension of the channel, is

$$\phi_N = \frac{1}{N} \sum_{i=1}^{N} \phi(u_i) \quad . \tag{3.9}$$

As previously mentioned, block coding arises in consideration of actual coding techniques to minimize distortion and increase reliability in a communication system. One naive approach to encoding might be to just repeat each scalar source symbol many times through the channel. In other words, for each source symbol x, construct a vector u whose components are x's. As the number of repetitions increases, the dimension N of u increases for a fixed dimension n of x. In the limit, this scheme will drive the probability of error to zero [4], but will pay a price. As N increases, the channel is taking longer and

longer to transmit the same amount of information being emitted from the source. Therefore, the source rate is decreasing relative to the channel rate. This means that the channel is being used inefficiently, since the rate at which it can handle inputs is much larger than the rate at which information is being produced. This tradeoff of rate and reliability was thought to be the best one could do, until Shannon came along. His theorem says that one can do much better; that is, for any fixed rate R, the probability of error can be driven to zero in the limit (and thus minimize distortion) by simultaneously increasing N and n and choosing clever encoders and decoders. This is the really crucial point of Shannon's theorem. Thus, both methods of block coding, i.e., repeating and the coding scheme referred to in Shannon theorem, are limiting results. As N and/or n increases to ∞, so does the delay T, the time it takes to emit one n-vector from the source or transmit one N-vector through the channel. In the case of repeating, this delay T is incurred every time a source symbol is emitted. However, in Shannon's theorem, the delay is incurred just once, at the beginning, when the first n-vector is emitted. Then the source and channel are matched synchronously, so that while a source vector is being produced, the previous source vector is simultaneously being sent through the channel. It takes T seconds to accomplish both these tasks, so that no more delay is incurred. In practice, the initial delay is not significant, relative to the entire time the communication system is in operation.

We are now ready to state the major result of classical information theory.

Shannon's Coding Theorem: Suppose a source and channel with α and β specified are given. If

$$R(\beta) \leq C(\alpha) \quad , \tag{3.10}$$

then for arbitrary $\epsilon_1 > 0$ and $\epsilon_2 > 0$, there exists a T sufficiently large and an encoder-decoder pair such that average cost satisfies $E[\phi] \leq \alpha + \epsilon_1$ and average distortion satisfies $E[D] \leq \beta + \epsilon_2$.

Proof: See [1] and [4].

Thus, $R(\beta)$ and $C(\alpha)$ are not exactly the source rate and channel capacity, respectively, but are approximations which become more exact as the delay T becomes large.

Converse to the Coding Theorem: If

$$R(\beta) > C(\alpha) \quad , \tag{3.11}$$

then there does not exist an encoder-decoder pair such that $E[\phi] = \alpha$ and $E[D] = \beta$.

Proof: See [1], [5], and [14].

In other words, (3.10) is the best we can do, for if (3.11) holds, then even in the limit, average distortion β cannot be attained. Another way of stating the converse is that if $E[D] = \beta$ can be attained (approximately) at a cost $E[\phi] = \alpha$, then α and β must satisfy (3.10). Then β^*, the solution to (3.10) with equality for given α, is a lower bound (called the Shannon bound) for attainable distortions. However, the really important result is the Coding Theorem itself, which states that β^* is actually attainable (in the limit as T increases).

3.2 Discussion

Several points can now be made about these results. First of all, as surprising as Shannon's theorem is, it has one major drawback: it is an existence theorem and, thus, does not provide a technique for actually constructing the encoder and decoder. The solutions to the optimization problems (3.4) and (3.3) for $R(\beta)$ and $C(\alpha)$, respectively, are not coders, but optimal probability density functions. They have limited usefulness in finding the optimal coders. For example, if an arbitrary coding scheme is constructed, its densities can be computed and compared against the solutions to (3.3) and (3.4) to see if the scheme is optimal. However, this seems to be about as far as one can go using only the Coding Theorem.

When we consider the rate distortion function together with the channel to find the minimum distortion, we can re-express $R(\beta)$ as (see [2])

$$\beta^* = \inf_{p(v|x)} E[D(x,v)] \quad \text{s.t.} \quad \rho_s I(x,v) \leq C(\alpha) \quad . \tag{3.12}$$

This formulation is appealing because it seems more natural to minimize distortion rather than rate. Suppose $p^*(v|x)$ is the probability density that attains β^*, and $I^*(x,v)$ is the corresponding mutual information evaluated with $p^*(v|x)$. Then

$$R(\beta^*) = \rho_s I^*(x,v) = C(\alpha) \quad . \tag{3.13}$$

This illustrates the close connection between the Coding Theorem and minimizing distortion.

4. Optimal Payoff and Strategies

Now that the fundamentals of Shannon theory have been established, we can return to the five team theory questions posed in Section 2. This section will address the first two questions. First of all, since the Coding Theorem holds only when infinite sequences of source and channel symbols are allowed, we must modify the team formulation to account for vectors x, u, etc. of arbitrarily large dimension. Then the team problem becomes

$$\min_{\gamma_1, \gamma_2} \quad J = E\left[\lim_{n\to\infty} D_n\right] \quad \text{s.t.} \quad E\left[\lim_{N\to\infty} \phi_N\right] \leq \alpha \,, \tag{3.14}$$

where γ_1 and γ_2 are now mappings between infinite vectors.* Since the objective is to minimize distortion, *the optimal value of* J, call it J^*, is just β^*, which satisfies

$$R(\beta^*) = C(\alpha) \,,$$

since $R(\beta)$ is a decreasing function of β bounded from above by $C(\alpha)$. Therefore, Shannon's theorem immediately gives us the optimal payoff for the team problem (3.14), and so answers Question 2 in the limit as $n, N \to \infty$ (or, equivalently, $T \to \infty$).

As a graphical interpretation of $J^* = \beta^*$, recall Example 3.1, a discrete, binary source. Suppose we turn the problem around and ask the following question:

---

*Wyner [14] formulated the problem this way, but not in the context of team theory.

Question: If we want to send data through a channel of capacity C (for fixed α) with distortion no more than β, what is the maximum possible rate at the source.

Answer: Set $R(\beta) = C$ and solve for $\rho_s(\beta)$.

From (3.5), this gives

$$\rho_s = \frac{C}{1-h(\beta)} \quad . \tag{3.15}$$

A plot of ρ_s vs. β is shown in Figure 3.6. It shows that when distortion is allowed, the data rate will be faster than the channel capacity, a somewhat *nonintuitive* result, since capacity is thought of as a maximum rate.

There is another way of getting this curve which illustrates the idea of (3.10). First plot $R(\beta)$ as a function of β for different values of ρ_s (see the solid curves in Figure 3.7). Then draw the line C. Shannon's result that $R(\beta) \leq C$ defines the forbidden and attainable regions. The best bound, namely, the minimum attainable β for a given ρ_s, is indicated by the circles, the points at which $R(\beta) = C$. Matching these optimal β's with their corresponding ρ_s's, we get the dotted curve, which is precisely the same curve as in Figure 3.6. Therefore, given the entire communication system, the points on the curve $\rho_s(\beta)$ are attained in three steps:

1. Fix ρ_c and solve for $C(\alpha) = C$ in (3.3).
2. Fix ρ_s and solve for $R(\beta)$ in (3.4).
3. Set $R(\beta) = C$ and solve for $\rho_s(\beta)$.

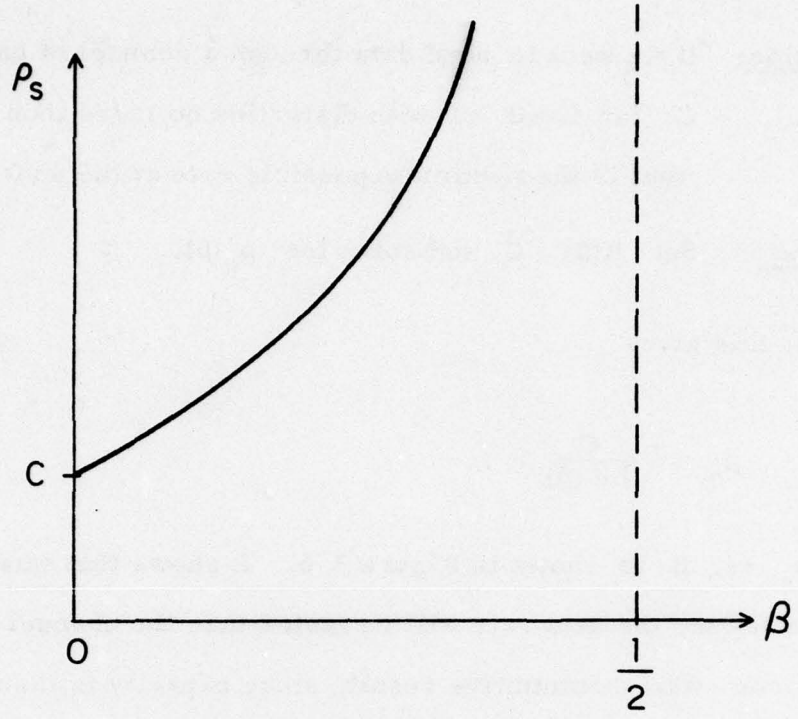


FIG. 3.6 $\rho_s(\beta)$ FOR BINARY SOURCE

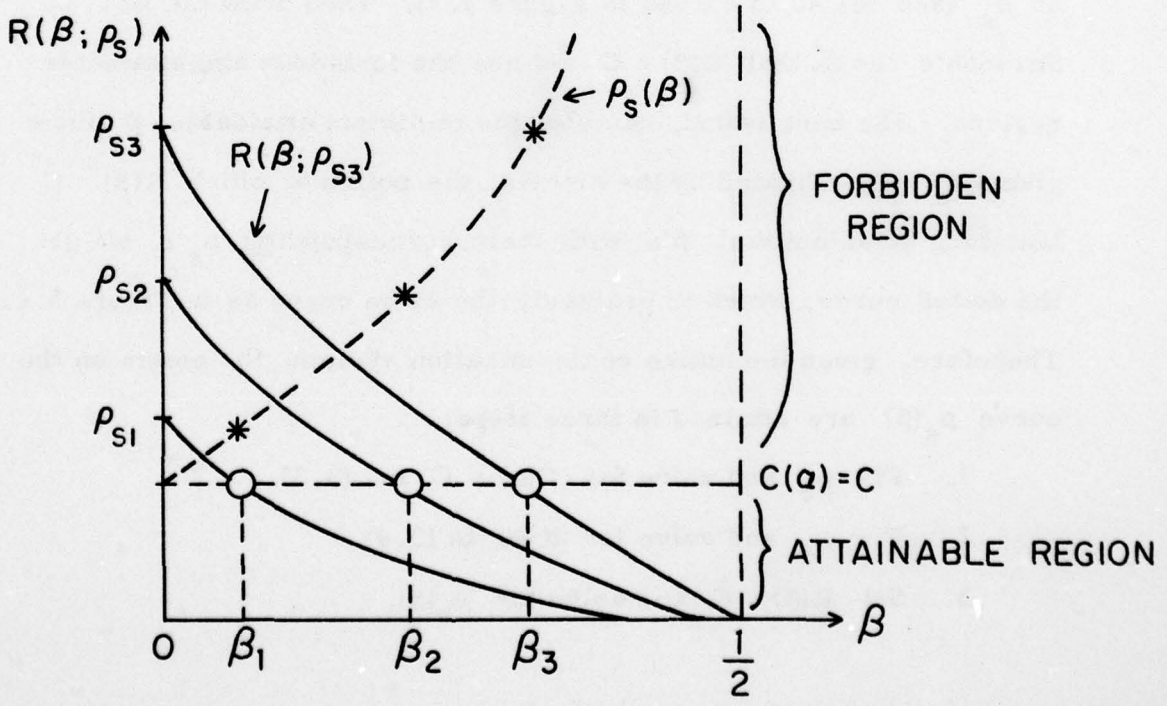


FIG. 3.7 OTHER DERIVATION OF $\rho_s(\beta)$ FOR BINARY SOURCE

If we solve (3.12) instead of (3.4), then we get (3.13) and can skip step 3 entirely.

This procedure for finding the minimum distortion can be thought of as a kind of "separation principle." The two-person team problem with dynamic information is replaced by two one-person control problems ((3.3) and (3.4)) with static information. Whittle and Rudge recognized this when they said: "Control and communication are both required but the controls operate on separate parts of the system, so that joint control is not required" (see [10], p. 366). However, as Witsenhausen [12] noted, this procedure is still not general enough to be applied to other team problems.

For fixed ρ_s, if the team problem could be solved, then it would yield $\beta(\rho_s)$. Inverting this function yields $\rho_s(\beta)$, precisely the same curve as in Figure 3.6 and the dotted curve in Figure 3.7. Thus, the condition $R(\beta) \leq C(\alpha)$ is buried inside the team formulation. It is not yet clear whether this inequality can be derived from the viewpoint of team theory alone. However, it can be derived from team theory together with rate distortion theory, which yields the solid curves in Figure 3.7. For example, in Figure 3.7, suppose that the rate curves $R(\beta)$ and $\rho_s(\beta)$ are shown, but the line $C(\alpha)$ is not. We will derive this line in another way, without reference to $R(\beta) \leq C(\alpha)$. Consider the points where a vertical line drawn through β_i intersects $R(\beta;\rho_{si})$ (let $P_i \triangleq R(\beta_i;\rho_{si})$, $i = 1, 2, 3$). The interpretation of β_i from $\rho_s(\beta)$ is that, given ρ_{si}, it is the minimum attainable distortion. Thus, for β such that $R(\beta;\rho_{si}) \leq P_i$, β is attainable, and for $R(\beta;\rho_{si}) > P_i$, β is not attainable. If we do this for all i, then we will find that the

P_i's are all equal; that is, the points of intersection lie on a horizontal line. Let $P_1 = P_2 = \cdots \triangleq C$. Then we immediately have the condition $R(\beta) \leq C$ defining the attainable region.

Since the solutions of (3.3) and (3.4) yield only the optimal probabilities $p(u)$ and $p(v|x)$, the optimal encoder and decoder are still not known. In fact, $C(\alpha)$ and $R(\beta)$ are computed on a "per-symbol basis," that is, with regard to a single input-output pair, whereas the Coding Theorem is a statement about transmission of information when there are infinite sequences. Thus, in the language of decision and control theory, Shannon's theorem provides the optimal payoff (in the limit as $T \to \infty$) but not the optimal strategies. It is purely an existence theorem. Its nonconstructive nature has frustrated information theorists to this day. Therefore, *Question 1 cannot, in general, be answered by Shannon theory.*

5. Real-Time Information Theory

5.1 Introduction

In this section, the problem of suboptimal strategies (Question 3) will be raised in the context of a new approach toward solving communication problems. In the previous two approaches discussed in Section 3.2--simple repeating and the cleverer coding scheme whose existence is proved by Shannon's theorem--it was assumed that sequences could be infinitely long, and thus incur an infinite delay. If the dimension of x is large, then the encoder must wait for the entire vector x before it starts to code. If the dimension of u is

large, the decoder must wait for the entire vector u to go through the channel, symbol by symbol. A third approach, which is the one to be discussed in the rest of this chapter, is what we call "real-time information theory." In this approach, both N and n are fixed. That is, we consider block codes with a fixed block length. This situation might occur if the receiver is another DM who must make decisions in real time, i.e., without arbitrary delay. The team problem (3.14) now becomes

$$\min_{\gamma_1, \gamma_2} J = E[D_n] \quad \text{s.t.} \quad E[\phi_N] \leq \alpha \quad , \tag{3.16}$$

where γ_1 and γ_2 are mappings between finite vectors. With this extra restriction of fixed length, the optimal encoder and decoder in (3.16) may not attain the Shannon bound in (3.14). That is, they are suboptimal in the infinite delay problem. However, the formulation in (3.16) is actually closer to traditional team theory, which does not deal with infinite or arbitrary delays. This assumption of fixed dimensions also brings us closer to the Spence problem, which can now be looked upon as a NZS version of real-time information theory where $n = N = 1$. Since dimensionality is at issue, Question 4 will also be answered, which completes our list of team theory questions.

The Shannon bound is the best we can do if n and N are allowed to become arbitrarily large; that is, the admissible strategy space contains mappings between vectors of arbitrary lengths. The mappings between vectors of fixed length n and N constitute a subset of this space. Since its strategy space is more restricted than in the

AD-A045 111 HARVARD UNIV CAMBRIDGE MASS DIV OF ENGINEERING AND --ETC F/G 12/2
INFORMATION AND SIGNALING IN DECENTRALIZED DECISION PROBLEMS.(U)
SEP 77 M P KASTNER
N00014-75-C-0648

UNCLASSIFIED TR-669 NL

2 OF 2
AD A045111

END
DATE FILMED
11-77
DDC

Shannon problem, the real-time team problem for fixed n and N cannot have a distortion less than β^*. Thus, β^* is a lower bound for the real-time problem. Witsenhausen [12] also pointed this out, and noted that the bound may be quite loose. He refers to a paper by Ziv and Zakai [15] that proposes a way to find tighter bounds by replacing "log" by some other convex function in the definition of an information measure. However, these bounds are, in general, difficult to compute.

In the next section, we will investigate a particular example of a communication system and will show under what circumstances linear strategies for fixed n and N attain the Shannon bound. When this happens, the strategies are, therefore, team optimal. When this does not happen, the performance of the suboptimal linear strategies can be compared to β^*. If the performance is close to β^*, then the easy-to-implement linear strategies might be desirable.

5.2 Linear vs. Nonlinear Strategies

In order to provide a basis for comparison of optimal and suboptimal strategies, we will assume that, for all examples discussed in this section, the communication system in question has a Gaussian source and channel, as described in Examples 3.2 and 3.3 but with variance (x) = 1. Then for fixed α, the minimum distortion β^*, derived from equating $R(\beta)$ from (3.6) and $C(\alpha)$ from (3.7), is

$$\beta^*(k) = \left(\frac{\alpha + \sigma^2}{\sigma^2} \right)^{-k} \qquad (3.17)$$

where

$$k = \frac{\rho_c}{\rho_s} \quad . \tag{3.18}$$

Then D_n and ϕ_N from (3.8) and (3.9) become

$$D_n = \frac{1}{n} \sum_{i=1}^{n} (x_i - v_i)^2 \tag{3.19}$$

$$\phi_N = \frac{1}{N} \sum_{i=1}^{N} u_i^2 \quad . \tag{3.20}$$

First consider the special case of $n = N = 1$. Then the source and channel are matched synchronously, so that $\rho_c = \rho_s$, or $k = 1$. The team formulation in (3.16) reduces to the simple form

$$\min_{\gamma_1, \gamma_2} \quad J = E[(\gamma_2 - x)^2] \quad \text{s.t.} \quad E\,\gamma_1^2 \leq \alpha \quad . \tag{3.21}$$

Since J is the same as J_2 from the Spence problem, and the constraint does not depend on γ_2, and the first order conditions for the unconstrained team optimal are the same as the first order conditions for a Nash equilibrium [7], then the optimal decoder is the same as in the Spence problem, namely,

$$\hat{x} = v = \gamma_2(y) = E_{/y}(x) \quad . \tag{3.22}$$

We cannot evaluate this conditional mean until we specify γ_1, since $y = \gamma_1(x) + \epsilon$. Suppose we let γ_1 be linear; that is,

$$u = \gamma_1(x) = ax \quad , \tag{3.23}$$

where "a" is a scalar. Since x and ϵ are Gaussian random variables, γ_2 is also linear, and (3.22) becomes

$$\hat{x} = v = \gamma_2(y) = \frac{a}{a^2+\sigma^2} y \quad , \tag{3.24}$$

and the constraint in (3.21) becomes

$$a^2 \leq \alpha \quad . \tag{3.25}$$

For $a = \sqrt{\alpha}$, the constraint is satisfied and

$$J = E\left[\left(\frac{\alpha}{\alpha+\sigma^2} - 1\right) x + \frac{\sqrt{\alpha}}{\alpha+\sigma^2} \epsilon \right]^2$$

$$= \frac{\sigma^2}{\alpha+\sigma^2} = \beta^*(1) \; ! \tag{3.26}$$

This linear scheme attains the Shannon bound, so that we immediately have the solution to the team problem in (3.21). Therefore, for the special case of $N = n = 1$, the linear strategies

$$\gamma_1(x) = \sqrt{\alpha}\, x \tag{3.27a}$$

$$\gamma_2(y) = \frac{\sqrt{\alpha}}{\alpha + \sigma^2} y = \hat{x} \tag{3.27b}$$

are optimal. This result is at first very surprising, because, as mentioned earlier, Witsenhausen [11] showed that for a similar team problem with signaling, the optimal linear solution was not the team optimal. However, the result (3.27) was also noted by Witsenhausen in a later paper [12], by Gallager [5], and by Whittle and Rudge [10].

For other cases of fixed n and N besides $n = N = 1$, linear strategies could again be tried and compared against $\beta^*(k)$. To make the source and channel synchronously compatible, choose n and N such that

$$\frac{N}{n} = \frac{\rho_c}{\rho_s} = k \ , \tag{3.28}$$

as described in the discussion of block coding in Section 3.1. The communication system now considered will involve an n-dimensional memoryless Gaussian source x with zero mean and covariance I_n (n-dimensional identity matrix), and an N-dimensional additive Gaussian channel whose noise ϵ has zero mean and covariance $\sigma^2 I_N$ and is independent from x. For the encoder, a linear strategy means

$$u = Hx \ , \tag{3.29}$$

where H is an $N \times n$ matrix. It will be assumed to be of maximal rank, since this is required in the proof of Theorem 3.1 below. This assumption has the interpretation in equation (3.29) of requiring the components of u to be uncorrelated. Since we know from Shannon theory that the Shannon bound is attained when the inputs u to the channel are uncorrelated (see [5]), H having maximal rank is a reasonable assumption.

From (3.22), v can also be expressed as a function of H. The version of the team problem in (3.16) with D_n and ϕ_N as in (3.19) and (3.20), respectively, now becomes:

$$\min_{H \in \mathcal{H}} J = E\left[\frac{1}{n} \operatorname{tr}(x - v(H))(x - v(H))^T\right] \text{ s.t. } E\left[\frac{1}{N} \operatorname{tr} H x x^T H^T\right] \leq \alpha , \tag{3.30}$$

where "tr" stands for trace and superscript "T" for transpose, and $\mathcal{H}$ is the set of $N \times n$ matrices of maximal rank.

THEOREM 3.1. Let H^* be the optimal solution to (3.30). Then for $k \geq 1$

$$J(H^*; k) = \frac{\sigma^2}{\alpha k + \sigma^2} . \tag{3.31}$$

Proof. See Appendix III-A.

In the proof of Theorem 3.1, H^* is derived in terms of its eigenvalues, not the matrix itself. However, as will now be shown, a particular H^*, with a simple interpretation, can be found. For $k \geq 1$, that is, the channel dimension is greater than or equal to the source dimension, a particular linear encoder is the one which corresponds to repeating; that is

$$\gamma_1 : \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \longrightarrow u = \begin{bmatrix} u_1 \\ \vdots \\ u_N \end{bmatrix} = \sqrt{\alpha} \begin{bmatrix} x_1 \\ \vdots \\ x_1 \end{bmatrix} \Big\} k \text{ times} \quad \begin{bmatrix} x_2 \\ \vdots \\ x_2 \end{bmatrix} \Big\} k \quad \begin{bmatrix} x_n \\ \vdots \\ x_n \end{bmatrix} \Big\} k$$

where $N/n = k$. The particular H that corresponds to this is (call it $\hat{H}$):

$$\hat{H} = \sqrt{\alpha} \cdot \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & & & & \\ \vdots & & & & O \\ 1 & 0 & & \cdots & \\ 0 & 1 & 0 & \cdots & \\ \vdots & \vdots & & & \\ 0 & 1 & 0 & \cdots & \\ & & & \ddots & \\ 0 & \cdots & & 0 & 1 \\ & & & & 1 \\ & O & & & \vdots \\ & & & & 1 \end{bmatrix} \begin{matrix} \\ \\ \} k \\ \\ \} k \\ \\ \\ \} k \end{matrix} \quad kn = N \qquad (3.32)$$

COROLLARY 3.1. For $k \geq 1$, $J(\hat{H}) = J(H^*)$.

Proof. See Appendix III-A.

Therefore, for $k \geq 1$, repeating is as good as the best linear encoder.

An immediate consequence of Theorem 3.1 is

COROLLARY 3.2. For the communication system described in Section 5.1, linear encoders and decoders are optimal if the source dimension n and channel dimension N are equal.*

*Whittle and Rudge [10] prove a more general result for the case of channels with memory.

Proof. If $n = N$, then $k = 1$ and $J(H^*; k = 1) = \beta^*(1)$. Figure 3.8 shows how $J(H^*)$ deviates from β^* as k increases. Q.E.D.

Since β^* is calculated assuming infinite sequences, but $J(H^*)$ is not, the converse of Corollary 3.2 cannot automatically be asserted. It may be true that if the dimensions n and N are fixed, linear strategies are the best we can do. However, in Section 5.3 counter-examples for $k > 1$ and $k < 1$ will be described, where certain nonlinear coders give lower distortion than the best linear ones.

Before presenting the counterexamples, we first give the heuristic interpretation as to why linear is best for $N = n$ but not necessarily for $N \neq n$ that was first proposed by Shannon [9] and later by Wozencraft and Jacobs [13]. Consider Figure 3.9 for $n = N = 2$ and Figure 3.10 for $n = 1$, $N = 2$. Figure 3.9 illustrates the linear case. The idea here is that a linear transformation maps the entire space of x's (R^2) to the entire space of u's (R^2); that is, it fills the u-space. To understand the significance of this, we must compare it with Figure 3.10. First we perform a transformation on the Gaussian random variable x so that it falls within a finite interval. This simplifies the explanation and is an important step in one of the counterexamples. Now, Corollary 3.1 says that the best linear transformation on x is as good as just repeating x twice, which implies that the optimal linear coder maps the finite x interval to the diagonal $u_1 = u_2 = x$ in the u-space. However, this does not take advantage of the higher dimensionality of u; that is, it does not "fill the space." A transformation that results in a curve that fills the

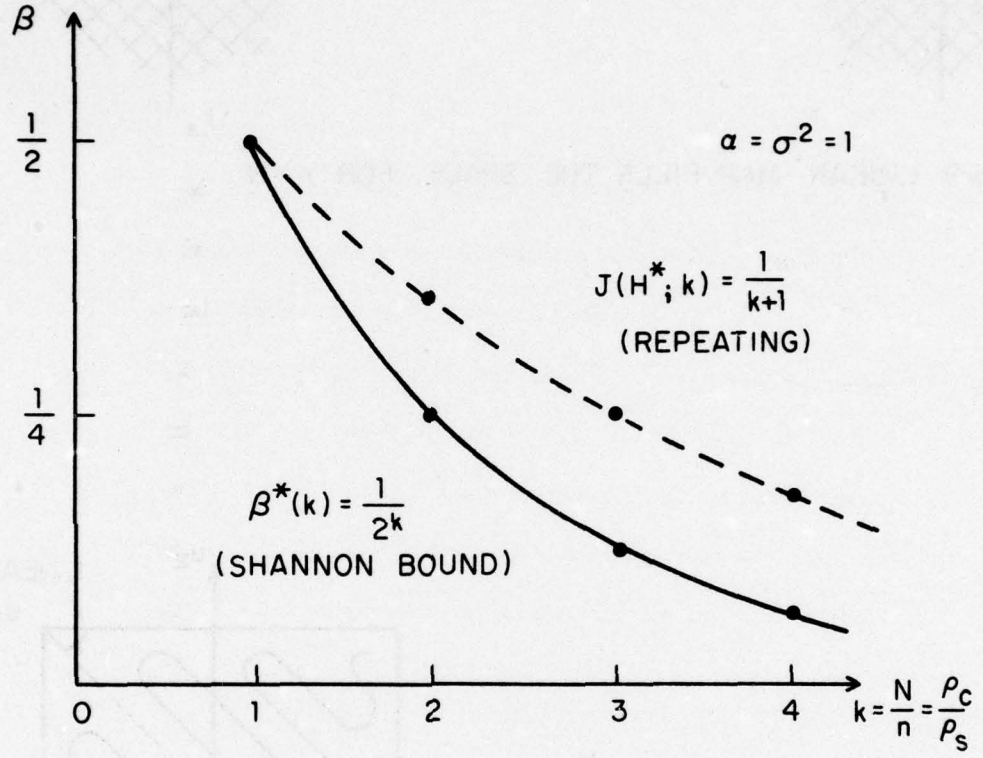


FIG. 3.8 COMPARISON OF SHANNON BOUND AND REPEATING

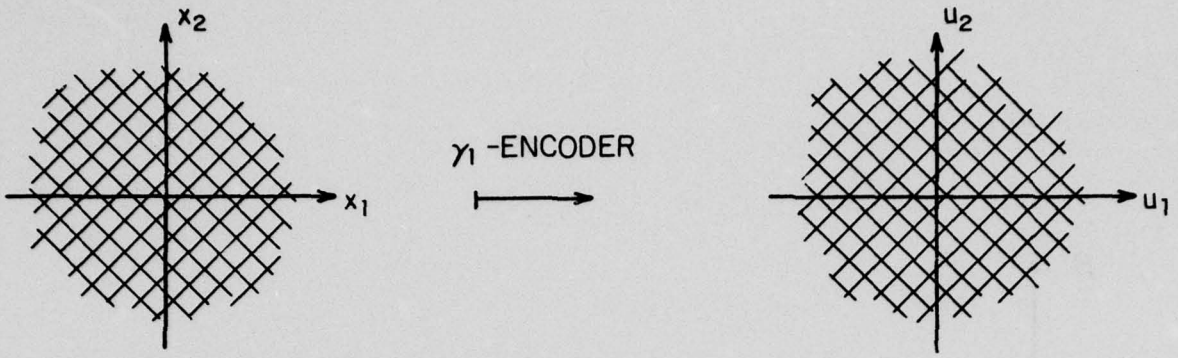


FIG. 3.9 LINEAR MAP FILLS THE SPACE FOR $n = N$

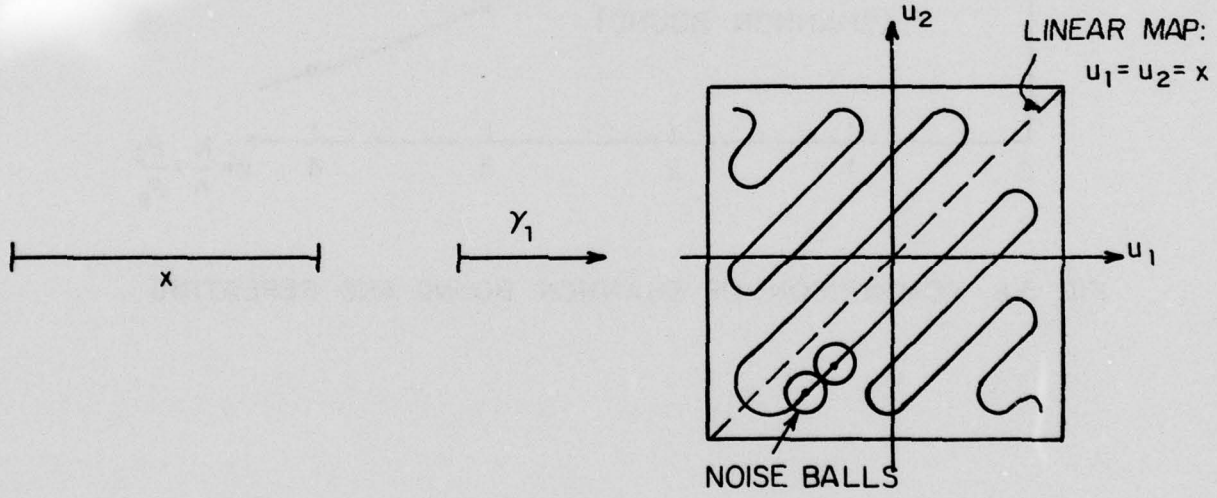


FIG. 3.10 STRETCHED CURVE BETTER THAN LINEAR FOR $n \neq N$

space more than the diagonal is better than linear, as illustrated by the twisting curve in Figure 3.10.* Now, when the signal (as represented by the curve) goes through the channel, it is corrupted by noise. The advantage of the longer curve is that we can pack in more little "noise balls," assuming that the variance σ^2 of the noise is very small in order to prevent accidentally jumping to the wrong part of the curve when the noise is added.** In fact, Shannon points out that there is a threshold effect where the increased benefits by extending the curve are outweighed by the greater chance of committing a large error. Now, define the "stretch factor" S (see [13]) as:

$$S = \frac{\text{change in length along curve}}{\text{change in } x} .$$

If S is constant all along the curve and the noise is small, then, locally within the balls, the curve looks linear. If we straighten out the curve and compress it to fit in the original interval in x, then we have also compressed the noise balls. The net effect is that we have reduced the noise for the whole system, so that we get a lower distortion than linear.***

5.3 Counterexamples

We now describe the counterexamples that show that linear strategies are not necessarily optimal when n and N are fixed, and

---

*Shannon [9] calls this idea the "snake-in-the-box."

**Although Gaussian noise extends beyond the boundaries of the noise balls, almost all of the probability density falls within a ball of radius 3σ. Thus, packing in balls captures the conceptual idea.

***In [13] this is called "twisted modulation."

$k \neq 1$. These examples utilize the "stretched curve" idea to construct nonlinear encoders and decoders that are better than linear. For simplicity, we assume that the random variable are normalized so that $\alpha = 1$.

COUNTEREXAMPLE 3.1.* $k > 1$.

Let $n = 1$ and $N = 2$, so that $x \sim N(0, I_1)$ and $\epsilon \sim N(0, \sigma^2 I_2)$. Divide x into four regions such that the probability density of each region is 1/4 (see Figure 3.11); that is,

$$\int_0^A p(x)\, dx = \int_A^\infty p(x)\, dx = \frac{1}{4} \quad \text{for} \quad p(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$$

$$\Rightarrow A \approx .67.$$

Since $N = 2$, the encoder must take x to some two-dimensional vector

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} .$$

The particular encoder used in this example is to let u_1 represent the region x is from, and let u_2 be a linear transformation of x in a stretched out version of this region. Figure 3.11 gives a graphical interpretation of this scheme. More precisely, let $r(x)$ = region number of x. Then it can be verified that u_2 can be expressed as

$$u_2 = B\left(\frac{2x}{A} + 5 - 2\, r(x)\right), \tag{3.33}$$

*See Appendix III-B for details.

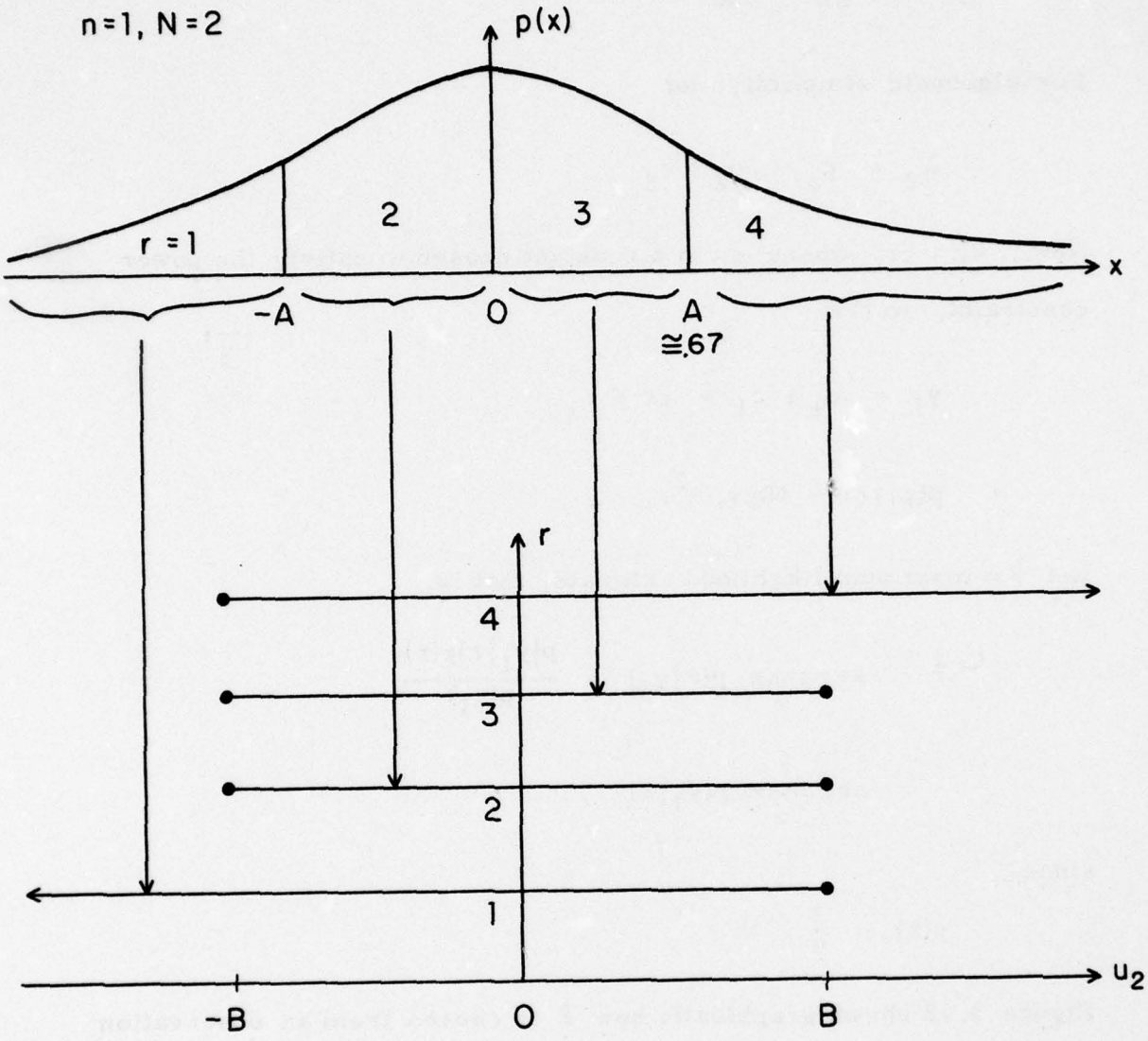


FIG. 3.11 STRETCHED MAPPING OF x TO TWO DIMENSIONS

(see Figure 3.11 for illustration of B) and the stretch factor S is constant for all x within a given region:

$$S = \frac{\Delta u_2}{\Delta x} = \frac{2B}{A} \qquad \text{for fixed } r.$$

For algebraic simplicity, let

$$\hat{u}_2 = y_2 = u_2 + \epsilon_2 \quad .$$

Now, $u_1 = cr$, where c is a constant chosen to satisfy the power constraint, so that

$$y_1 = u_1 + \epsilon_1 = cr + \epsilon_1$$

$$\Rightarrow \quad p(y_1|r) \sim N(cr, \sigma^2) \quad .$$

Let $\hat{r}$ = maximum likelihood estimate, that is,

$$\hat{r} = \arg\max_r \; p(r|y_1) = \frac{p(y_1|r)p(r)}{p(y_1)}$$

$$= \arg\max_r \; p(y_1|r) \quad ,$$

since

$$p(r) = \frac{1}{4} \quad \forall \; r \quad .$$

Figure 3.12 shows graphically how $\hat{r}$ is chosen from an observation of y_1. Finally, let

$$v = \hat{x} = \frac{A}{2}\left(\frac{\hat{u}_2}{B} - 5 + 2\hat{r}\right)$$

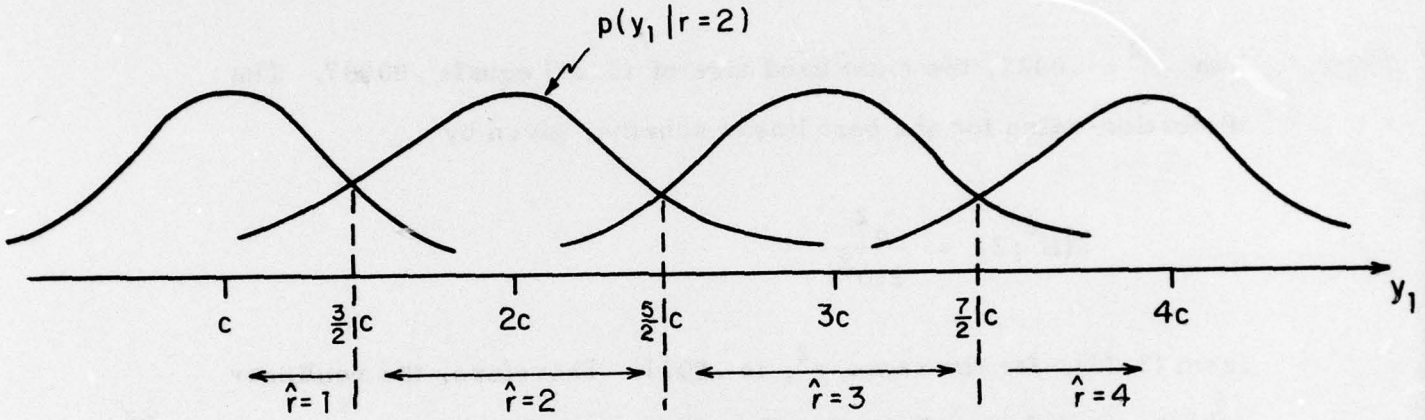


FIG. 3.12 ESTIMATING r

(just invert (3.33)). Then it can be shown that the expected distortion can be bounded from above:

$$E[(x-\hat{x})^2] \leq \frac{1}{6}\sigma^2 + 6F\left(-\frac{.18}{\sigma}\right) \quad , \tag{3.34}$$

where

$$F(y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx \, .$$

For $\sigma^2 = .0022$, the right hand side of (3.34) equals .00097. The distortion value for the best linear scheme, given by

$$J(H^*;2) = \frac{\sigma^2}{2+\sigma^2}$$

from (3.31), for the same σ^2, is .0011. Therefore, the nonlinear scheme gives lower distortion than just repeating. Since the error function F decreases very rapidly as σ decreases, the nonlinear scheme becomes even better with smaller σ. For example, for $\sigma^2 = .001$, (3.34) equals .00017, and (3.31) equals .00052.

COUNTEREXAMPLE 3.2.* $k < 1$

Let $n = 2$ and $N = 1$. It can be shown that an optimal linear encoder is $u = (x_1 + x_2)/\sqrt{2}$, with expected distortion

$$J_L^*(\sigma^2) = \frac{\frac{1}{2} + \sigma^2}{1 + \sigma^2} \, .$$

*See Appendix III-C for details.

(In general, linear distortion is

$$\frac{\frac{n-1}{n} + \sigma^2}{1 + \sigma^2} \quad .)$$

Graphically, this scheme amounts to projecting all points in the x_1, x_2-space to the diagonal $x_1 = x_2$. One way to "fill the space" better, suggested by Shannon [9], is to construct a real number u by alternating the digits of x_1 and x_2, that is, if

$$x_1 = \cdot\, a_1\, a_2\, a_3 \cdots$$

$$x_2 = \cdot\, b_1\, b_2\, b_3 \cdots$$

then

$$u = \cdot\, a_1\, b_1\, a_2\, b_2\, a_3\, b_3 \cdots \quad .$$

This nonlinear scheme fills the x_1, x_2-space much more than linear, but it is difficult to deal with analytically.

A simpler nonlinear scheme that fills the space more than linear is shown in Figure 3.13, where θ_1 and θ_2 are transformations f of x_1 and x_2, respectively, to the interval $[-1, 1]$. All points are mapped to the dotted lines in the following way:

$$a \le \theta_2 \le b \Rightarrow (\theta_1, \theta_2) \rightarrow (\theta_1, \tilde{\theta}_2) = \left(\theta_1, \frac{a+b}{2}\right) \quad ,$$

where

$$(a, b) \in \left\{\left(-1, -\frac{1}{2}\right), \left(-\frac{1}{2}, 0\right), \left(0, \frac{1}{2}\right), \left(\frac{1}{2}, 1\right)\right\} \quad .$$

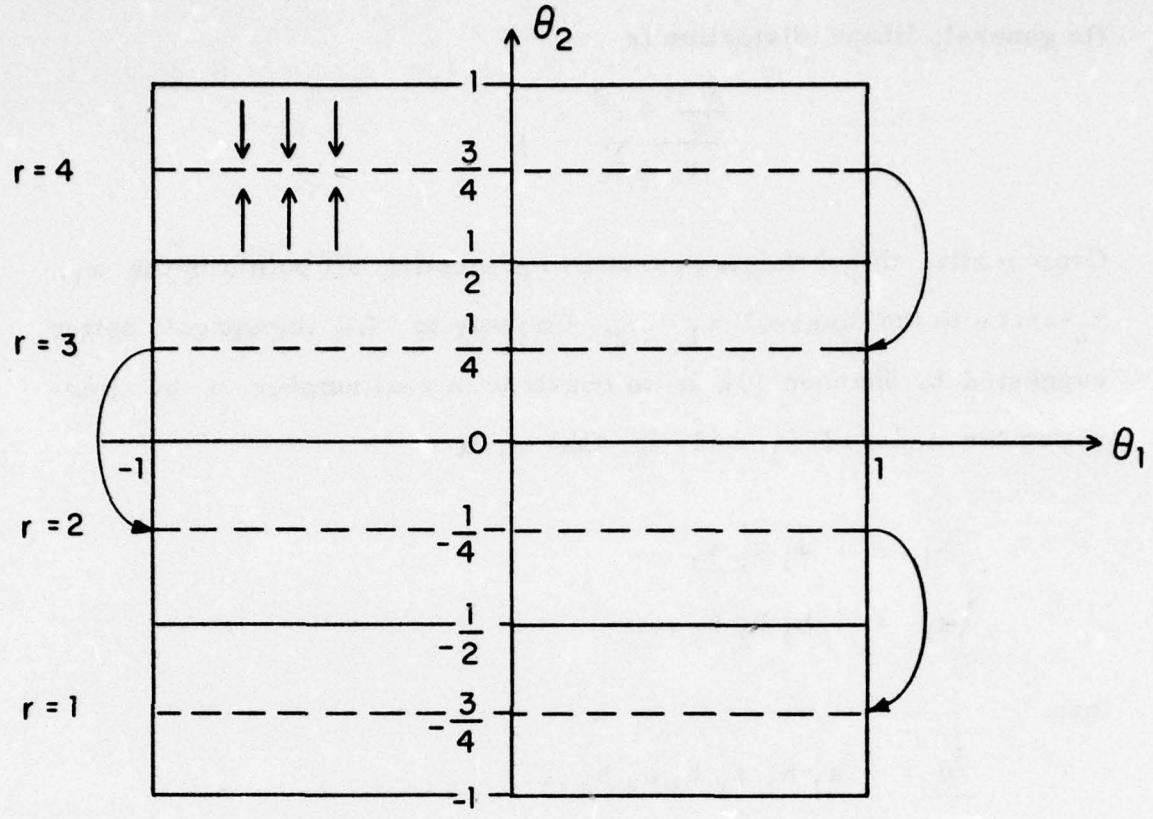


FIG. 3.13 TRANSFORMATION OF SQUARE TO DOTTED LINE

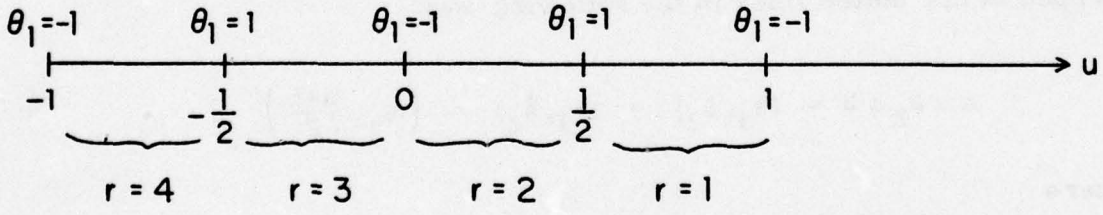


FIG. 3.14 TRANSFORMATION OF DOTTED LINE TO u

Let $\tilde{\theta}_2 = 3/4$ correspond to row number $r = 4$, $\tilde{\theta}_2 = 1/4$ to $r = 3$, etc. Straighten out the dotted line and compress it to fit into the interval $[-1, 1]$, and call the variable u, as shown in Figure 3.14. Then it can be shown that

$$u = \frac{1}{4}\left[(-1)^r \theta_1 + 5 - 2r\right] \quad . \tag{3.35}$$

Next u is sent through the channel. Let $\hat{u} = y = u + \epsilon$. Then let

$$\hat{r} = \begin{cases} 4 & \text{if} \quad \hat{u} \le -1 \\ 3 & \text{if} \quad -\frac{1}{2} \le \hat{u} \le 0 \\ 2 & \text{if} \quad 0 \le \hat{u} \le \frac{1}{2} \\ 1 & \text{if} \quad \frac{1}{2} \le \hat{u} \end{cases}$$

$$\hat{\theta}_1 = \begin{cases} (-1)^{\hat{r}}\,(4\hat{u} - 5 + 2\hat{r}), & -1 \le \hat{u} \le 1 \quad \text{(just invert (3.35))} \\ -1 \quad , & \hat{u} < -1 \\ 1 \quad , & \hat{u} > 1 \end{cases}$$

$$\Rightarrow \hat{\theta}_1 \in [-1, 1]$$

$$\hat{\theta}_2 = \frac{5 - 2\hat{r}}{4}$$

$$\hat{x}_1 = f^{-1}(\hat{\theta}_1)$$

$$\hat{x}_2 = f^{-1}(\hat{\theta}_2) \quad .$$

Let $f(x) = 2/\pi \tan^{-1} x$, and let expected distortion be defined as $J_{NL}(\sigma^2)$:

$$J_{NL}(\sigma^2) \triangleq \frac{1}{2} E[(x_1-\hat{x}_1)^2 + (x_2-\hat{x}_2)^2] \quad .$$

Then it can be shown that

$$J_{NL}(\sigma^2) \leq B(\sigma^2) \quad ,$$

where

$$B(\sigma^2) = \frac{1}{2} \left\{ \left[\frac{3}{2} \pi^2 16 \sigma^2 + .354 \right] (1-P_e) + \left[10\frac{1}{2} \pi^2 \right] P_e \right\} \qquad (3.36)$$

and $P_e = Pr[\hat{r} \neq r]$ = probability of error. P_e is a complicated expression because it involves the probability density of $\hat{u}$, a difficult thing to compute. However, P_e is a continuous increasing function of σ^2. If $\sigma^2 = 0$, then $P_e = 0$ and $B(0) = .177 = J_{NL}(0)$. If σ^2 is sufficiently small, then B will still be less than $J_L^*(\sigma^2)$, as shown in Figure 3.15. However, as $\sigma^2 \to \infty$, $P_e \to 1$, so that $B \to 5\frac{1}{4} \pi^2$. (The $B(\sigma^2)$ curve is qualitative and not based on numerical calculations.) Therefore, for sufficiently small σ^2, the nonlinear scheme is better than linear.

5.4 Asymptotic Effects

In Chapter II it was shown that the equilibrium solutions for Example 2.3 of the job market model exhibited threshold effects. Since the solutions to the $n = N$ case of the Shannon problem are known, we might ask whether they also exhibit threshold effects.

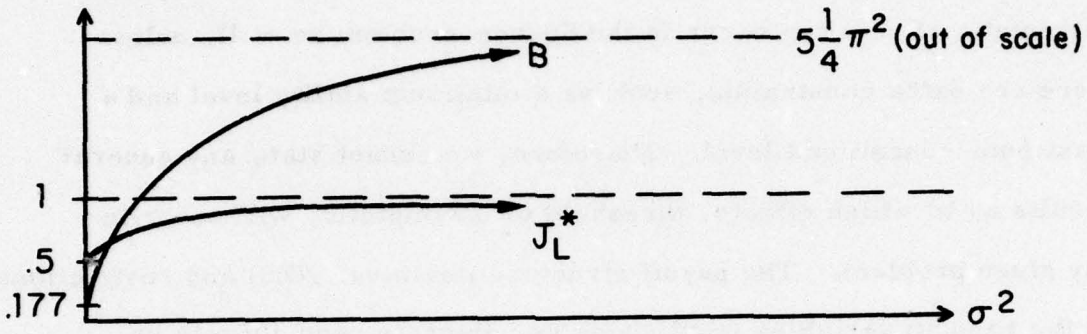


FIG. 3.15 OPTIMAL LINEAR PAYOFF VS. NONLINEAR BOUND

However, as expressions (3.27) and (3.26) for the $n = N = 1$ case indicate, asymptotic, not threshold, effects occur. That is, signaling ceases in the limit as the parameters α and σ^2 approach 0 or ∞. For example, both DMs still feel it worthwhile to signal even if signaling is very costly (tighter constraint on signal power, i.e., small α) or very noisy (large σ^2).

With no noise, $v = x$ as in Example 2.1, demonstrating that asymptotic effects can occur in the Spence problem as well, unless there are extra constraints, such as a minimum ability level and a maximum educational level. Therefore, we cannot state any general results as to which effects, threshold or asymptotic, will occur in any given problem. The payoff structure (team vs. NZS) and restrictions on the random variables (continuous vs. discrete, and infinite vs. finite range) are prime candidates as the factors which determine the type of parameter effects.

5.5 Summary

In general, real-time information theory solutions are sub-optimal as compared to "infinite-time" (Shannon) theory solutions. However, if the dimension n of a Gaussian source with a mean square error distortion function is equal to the dimension N of a memoryless Gaussian channel with a square cost function, the source and channel can be directly connected, with appropriate scaling of the channel inputs, so as to satisfy the power constraint. The distortion incurred is the best one can do, since it equals the Shannon bound. If

$N > n$ ($k = N/n > 1$), then repeating each source symbol k times is a simple suboptimal strategy. For small values of k, its performance is close to optimal. If $N < n$ ($k < 1$), Theorem 3.1 does not apply, so that little, in general, can be said. However, counterexamples with $k = 2$ and $1/2$ are described where nonlinear encoders are better than the best linear ones.

REFERENCES:

1. N. Abramson, Information Theory and Coding, New York: Mc Graw-Hill, 1963.

2. T. Berger, Rate Distortion Theory, New Jersey: Prentice-Hall, 1971.

3. A.E. Bryson and Y.C. Ho, Applied Optimal Control, Mass.: Ginn, 1969.

4. R. G. Gallager, "A Simple Derivation of the Coding Theorem and Some Applications," IEEE Trans. Inform. Theory, Vol. IT-11, Jan. 1965, pp. 3-18.

5. R. G. Gallager, Information Theory and Reliable Communication, New York: Wiley, 1968.

6. D. Middleton, Introduction to Statistical Communication Theory, New York: Mc Graw-Hill, 1960.

7. R. Radner, "Team Decision Problems," Ann. Math. Stat., Vol. 33, 1962, pp. 857-881.

8. C. E. Shannon, "A Mathematical Theory of Communication," Bell Sys. Tech. J., Vol. 27, 1948, pp. 379, 623.

9. C. E. Shannon, "Communication in the Presence of Noise," Proc. IRE, Vol. 37, Jan. 1949, pp. 10-21.

10. P. Whittle and J. F. Rudge, "Simultaneous Communication and Control," Adv. Appl. Prob., Vol. 8, 1976, pp. 365-384.

11. H. S. Witsenhausen, "A Counterexample in Stochastic Optimum Control," SIAM J. Control, Vol. 6, No. 1, 1968, pp. 131-147.

12. H. S. Witsenhausen, "The Intrinsic Model for Discrete Stochastic Control: Some Open Problems," Lecture Notes in Econ. and Math. Systems, Vol. 107, 1974, pp. 322-335.

13. J. M. Wozencraft and I. M. Jacobs, Principles of Communication Engineering, New York: Wiley, 1965.

14. A. D. Wyner, "Another Look at the Coding Theorem of Information Theory - A Tutorial," Proc. IEEE, Vol. 58, No. 6, June 1970, pp. 894-913.

15. J. Ziv and M. Zakai, "On Functionals Satisfying a Data-Processing Theorem," IEEE Trans. Inform. Theory, Vol. IT-19, No. 3, May 1973, pp. 275-283.

APPENDIX III-A

PROOFS FOR RESULTS ON OPTIMAL LINEAR CODERS

Before going on to prove Theorem 3.1, we digress to note that for the linear solution (3.27), $\hat{x} = v = \gamma_2^*(y)$ is the estimate of x that would be produced by a Kalman filter. Expected distortion $E[D(x, v)]$ is simply the error covariance, referred to as P in the control literature (see Bryson and Ho [3] for derivation of Kalman filter formulas). This fact makes it very easy to evaluate $\hat{x}$ and P from linear encoders and decoders for arbitrary values of N and n, using the standard formulas from Kalman filters for the special case where x is time-invariant. In general, if $x \sim N(\bar{x}, M)$, where M is the $n \times n$ covariance of x, $\epsilon \sim N(0, R)$, and $y = Hx + \epsilon$, H an $N \times n$ matrix, then from the Kalman filter

$$\hat{x} = \bar{x} + PH^T R^{-1}(y - H\bar{x}) \tag{3A.1}$$

$$P^{-1} = M^{-1} + H^T R^{-1} H \quad . \tag{3A.2}$$

In the special case of $n = N = 1$, it is easy to check that for our model, $H = \sqrt{\alpha}$, and $\hat{x}$ reduces to Equation (3.27b) and P to (3.26). If we restrict ourselves to linear encoders γ_1 that satisfy the power constraint with equality,* then the constraint

---

*We always assume equality in the constraint, because the rate of transmission increases with power. Thus, it is advantageous to use up all the power available.

$$E\left[\frac{1}{N}\sum_{i=1}^{N} u_i^2\right] = \alpha$$

becomes

$$\alpha = \frac{1}{N}\,\mathrm{tr}\; E\,uu^T$$

$$= \frac{1}{N}\,\mathrm{tr}\; E(Hxx^TH^T)$$

$$= \frac{1}{N}\,\mathrm{tr}\; (H^THExx^T)$$

$$= \frac{1}{N}\,\mathrm{tr}\; (H^TH) \quad . \tag{3A.3}$$

From (3A.2) and (3A.3), (3.30) becomes:

$$\min_{H\in\mathcal{H}} J(H) = \frac{1}{n}\,\mathrm{tr}\left(I_n + \frac{1}{\sigma^2}H^TH\right)^{-1} \quad \text{s.t.} \quad \frac{1}{N}\,\mathrm{tr}\;(H^TH) = \alpha, \tag{3A.4}$$

where $\mathcal{H}$ is the set of $N \times n$ matrices of maximal rank.

THEOREM 3.1. Let H^* be the optimal solution to (3.30). Then for $k \geq 1$

$$J(H^*;k) = \frac{\sigma^2}{\alpha k+\sigma^2} \tag{3.31}$$

Proof. Since $I_n > 0$ and $H^TH \geq 0$, then

$$I_n + \frac{1}{\sigma^2}H^TH > 0,$$

and from (3A.4)

$$J(H) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\lambda_i} \quad ,$$

where the λ_i are the eigenvalues of $I_n + 1/\sigma^2 H^T H$ and $\lambda_i > 0$ for all i. It is trivial to show that

$$\lambda_i = 1 + \mu_i, \qquad \forall\, i = 1, \ldots, n$$

where the μ_i are the eigenvalues of $1/\sigma^2 H^T H$. Then the constraint (3A. 3) becomes

$$\frac{\sigma^2}{N} \sum_{i=1}^{n} \mu_i = \alpha \quad .$$

But

$$\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} (1 + \mu_i) = n + \frac{N\alpha}{\sigma^2} \quad .$$

So the problem is now:

$$\min_{\{\lambda_i\}} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\lambda_i} \quad \text{s.t.} \quad \sum_{i=1}^{n} \lambda_i = n + \frac{N\alpha}{\sigma^2} \quad .$$

Let

$$L = \sum_{i=1}^{n} \frac{1}{\lambda_i} + r \left(\sum_{i=1}^{n} \lambda_i - n - \frac{N\alpha}{\sigma^2} \right) \quad ,$$

where r is the Lagrange multiplier.

$$\frac{\partial L}{\partial \lambda_i} = -\frac{1}{\lambda_i^2} + r = 0, \qquad i = 1, \ldots, n$$

$$\Rightarrow \quad \lambda_i^* = \frac{1}{\sqrt{r}} \quad \forall i \qquad \text{(positive square root because eigenvalues positive).}$$

In order for all λ_i^* to be equal, we must have all $u_i^* = \lambda_i^* - 1$ equal. This means we must have $H^T H > 0$. Since $k \geq 1$ and H is of maximal rank, then $H^T H$ is, in fact, positive definite (see [3], p. 444). (If $k < 1$, then $H^T H$ can only be assumed to be positive semi-definite.) Then

$$\sum_{i=1}^{n} \lambda_i^* = \frac{n}{\sqrt{r}} = n + \frac{N\alpha}{\sigma^2}$$

$$\Rightarrow \quad \sqrt{r} = \frac{n\sigma^2}{n\sigma^2 + N\alpha} = \frac{\sigma^2}{\sigma^2 + \alpha k}$$

$$\therefore \quad J(H^*) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\lambda_i^*} = \frac{1}{n} \sum_{i=1}^{n} \sqrt{r} = \sqrt{r}$$

$$= \frac{\sigma^2}{\sigma^2 + \alpha k} \quad .$$

Q. E. D.

COROLLARY 3.1. For $k \geq 1$, $J(\hat{H}) = J(H^*)$.

Proof. With $\hat{H}$ as in (3.32), then

$$J(\hat{H}) = \frac{1}{n} \operatorname{tr} P = \frac{1}{n} \operatorname{tr} \left(I_n + \frac{1}{\sigma^2} \hat{H}^T \hat{H} \right)^{-1}$$

$$= \frac{1}{n} \operatorname{tr} \left(I_n + \frac{1}{\sigma^2} \alpha k I_n \right)^{-1}$$

$$= \frac{\sigma^2}{\alpha k + \sigma^2} \quad . \qquad \text{Q. E. D.}$$

Therefore, within the class of linear encoders H, none gives lower distortion than $\hat{H}$.

APPENDIX III-B

COMPUTATIONAL DETAILS FOR COUNTEREXAMPLE 3.1

Referring to Figure 3.13, for

$$r = 1, \quad \text{let} \quad u_2 = B\left(\frac{2x}{A} + 3\right)$$

$$r = 2, \quad \text{let} \quad u_2 = B\left(\frac{2x}{A} + 1\right)$$

$$r = 3, \quad \text{let} \quad u_2 = B\left(\frac{2x}{A} - 1\right)$$

$$r = 4, \quad \text{let} \quad u_2 = B\left(\frac{2x}{A} - 3\right)$$

or

$$u_2 = B\left(\frac{2x}{A} + g(r(x))\right) \tag{3B.1}$$

where

$$g(r) = 5 - 2r \quad . \tag{3B.2}$$

Since $\alpha = 1$, the power constraint has been normalized to

$$\frac{1}{2} E(u_1^2 + u_2^2) = 1 \quad .^*$$

Let $u_1 = cr$ for some scalar c. Then a special case of the constraint is:

---

*As in Appendix III-A, we assume equality in the constraint.

$$1 = Eu_1^2 = c^2 Er^2 = \frac{15}{2} c^2$$

$$\Rightarrow \quad c = \sqrt{\frac{2}{15}} \cong .36 \quad .$$

Also

$$1 = Eu_2^2 = B^2 E\left[\frac{4x^2}{A^2} + \frac{4x}{A}\, g(r(x)) + g(r(x))^2\right]$$

$$= B^2\left[\frac{4}{A^2} + \frac{4}{A}\, E(xg) + 5\right] \quad ,$$

where

$$E(xg) = -2\left[\int_0^A x\, p(x)\, dx + 3\int_A^\infty x\, p(x)\, dx\right] \cong -2.08.$$

Therefore

$$1 = 1.5\, B^2,$$

or

$$B^2 = \frac{2}{3} \quad , \qquad B \cong .82 \quad . \tag{3B.3}$$

Figure 3.14 shows graphically how $\hat{r}$ is chosen from an observation of y_1. Transform y_1 to $\tilde{y}_1 = (y_1 - cr)/\sigma \sim N(0, 1)$. Then the cut-off points $(3/2)c$, $(5/2)c$, $(7/2)c$ in Figure 3.14 are transformed to $\pm c/2\sigma$. Let F be the error function

$$F(y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\, dx \quad .$$

Then for P_e = probability of error,

$$P_e \triangleq \Pr[\hat{r} \neq r] = \sum_{i=1}^{4} \Pr[\hat{r} \neq r \,|\, r=i] \Pr[r=i]$$

$$= \frac{1}{4}\left[F\left(-\frac{c}{2\sigma}\right) + 2F\left(-\frac{c}{2\sigma}\right) + 2F\left(-\frac{c}{2\sigma}\right) + F\left(-\frac{c}{2\sigma}\right)\right]$$

$$= \frac{3}{2} F\left(-\frac{c}{2\sigma}\right) \quad .$$

Note that as σ decreases, P_e decreases. Now, let

$$\hat{x} = \frac{A}{2}\left(\frac{\hat{u}_2}{B} - g(\hat{r}(x))\right)$$

(just invert (3B. 1)). Then

$$\hat{x} - x = \frac{A}{2}\left[\frac{1}{B}(u_2 + \epsilon_2) - g(\hat{r})\right] - x$$

$$= \frac{A}{2}(g(r) - g(\hat{r})) + \frac{A}{2B}\,\epsilon_2$$

$$E[(\hat{x}-x)^2] = \frac{A^2}{4B^2}\sigma^2 + \frac{A^2}{4} E[(g(r) - g(\hat{r}))^2]$$

$$= \frac{\sigma^2}{6} + A^2 E[(r-\hat{r})^2], \qquad \text{from (3B. 2) and (3B. 3)}$$

$$= \frac{\sigma^2}{6} + A^2 E[(r-\hat{r})^2 \mid \hat{r} \neq r]\, P_e$$

$$\leq \frac{\sigma^2}{6} + A^2(3)^2 P_e \;, \quad \text{since } \max |r-\hat{r}| = 4-1 = 3,$$

$$= \frac{1}{6}\sigma^2 + 6F\left(-\frac{.18}{\sigma}\right) \quad .$$

APPENDIX III-C

COMPUTATIONAL DETAILS FOR COUNTEREXAMPLE 3.2

A general linear scheme for this example is

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \rightarrow u = ax_1 + bx_2 \quad .$$

Power constraint:

$$E[u^2] = 1 = E[(ax_1 + bx_2)^2] = a^2 + b^2 \tag{3C.1}$$

Then

$$y = u + \epsilon$$

$$= ax_1 + bx_2 + \epsilon$$

$$= ax_1 + (bx_2 + \epsilon) \Rightarrow \hat{x}_1 = E_{/y}(x_1) = \frac{a}{a^2+b^2+\sigma^2} \; y$$

$$= bx_2 + (ax_1 + \epsilon) \Rightarrow \hat{x}_2 = E_{/y}(x_2) = \frac{b}{a^2+b^2+\sigma^2} \; y$$

$$E[\, D(x, v)\,] \triangleq \bar{D} = \frac{1}{2} E\,[(x_1-\hat{x}_1)^2 + (x_2-\hat{x}_2)^2]$$

$$= \frac{1}{2(a^2+b^2+\sigma^2)^2} \; [(a^2-1-\sigma^2)^2 + 2a^2b^2 + (b^2-1-\sigma^2)^2 + (a^2+b^2)\sigma^2] \quad .$$

Substituting in the constraint (3C. 1), this becomes

$$\bar{D} = \frac{1}{2(1+\sigma^2)^2} [2\sigma^4 + 3\sigma^2 + 1]$$

$$= \frac{\frac{1}{2} + \sigma^2}{1 + \sigma^2} \triangleq J_L^*(\sigma^2) \quad , \qquad (3C.2)$$

which is independent of a and b. Therefore, any a and b satisfying (3C. 1) will be an optimal linear solution.* For example, consider

$$a = b = \frac{1}{\sqrt{2}} \quad ,$$

so that

$$u = \frac{1}{\sqrt{2}} (x_1 + x_2) \quad .$$

For the nonlinear scheme described in Counterexample 3. 2, recall that $\theta = f(x)$. For the case of $f(x) = 2/\pi \tan^{-1} x$,

$$\hat{\theta}_2 = \pm \frac{3}{4} \Rightarrow \hat{x}_2 = \pm 2.4$$

$$\hat{\theta}_2 = \pm \frac{1}{4} \Rightarrow \hat{x}_2 = \pm .41$$

$$\theta_2 = \pm \frac{1}{2} \Rightarrow x_2 = \pm 1 \quad .$$

If $\sigma^2 = 0$ (no noise), then $\hat{\theta}_1 = \theta_1$ and $\hat{\theta}_2 = \tilde{\theta}_2$, so that

*That is, the constraint is rotationally invariant.

$$J_{NL}(\sigma^2=0) = \frac{1}{2} E[(x_1-\hat{x}_1)^2 + (x_2-\hat{x}_2)^2]$$

$$= \int_0^1 (x_2-.41)^2 p(x_2)\, dx_2 + \int_1^\infty (x_2-2.4)^2 p(x_2)\, dx_2$$

$$= .177 \quad .$$

From (3C.2), $J_L^*(0) = .5 > J_{NL}(0)$. Therefore, since the nonlinear coding gives lower distortion than the linear when there is no noise, it must do the same for sufficiently small noise. In fact, we can bound $J_{NL}(\sigma^2)$ for small σ^2.

Let

$$g(\theta_i) = g(f(x_i)) = x_i \qquad (\text{i.e., } g = f^{-1}) \quad . \tag{3C.3}$$

Then

$$g'(f(x_i))\, f'(x_i) = 1 \;, \quad \text{so that} \quad g'(f(x_i)) = \frac{1}{f'(x_i)} \quad . \tag{3C.4}$$

Let

$$\delta_1 = \hat{\theta}_1 - \theta_1 \quad .$$

Then

$$\delta_1^2 = (\hat{\theta}_1-\theta_1)^2 = \begin{cases} 16(\hat{u}-u)^2 = 16\epsilon^2 & \text{if } \hat{r} = r \\ \leq 4 & \text{if } \hat{r} \neq r \text{, since } \max|\hat{\theta}_1-\theta_1| = 1-(-1) = 2 \end{cases}$$

Let

$$\delta_2 = \hat{\theta}_2 - \theta_2 \quad .$$

Then

$$\delta_2^2 = \begin{cases} (\tilde{\theta}_2-\theta_2)^2 & \text{if } \hat{r}=r \text{ (or } E[(x_2-\hat{x}_2)^2 \mid \hat{r}=r] = .354 = 2J_{NL}(0)) \\ (\hat{\theta}_2-\theta_2)^2 \le \left(1\frac{3}{4}\right)^2 \approx 3 & \text{if } \hat{r}\neq r, \text{ since } \max|\hat{\theta}_2-\theta_2| = \frac{3}{4}-(-1) = 1\frac{3}{4} \end{cases}$$

$$\hat{x}_i = g(\hat{\theta}_i) = g(\theta_i) + g'(\theta_i)\delta_i + \frac{g''(\tau_i)}{2!}\,\delta_i^2 \quad ,$$

τ_i between θ_i and $\hat{\theta}_i$

$$\approx x_i + \frac{1}{f'(x_i)}\,\delta_i$$

from (3C.3) and (3C.4) (for very small δ_i, i.e., small noise*)

$$\therefore\ \hat{x}_i - x_i \approx \frac{1}{f'(x_i)}\,\delta_i$$

$$J_{NL}(\sigma^2) = \frac{1}{2}\,E[(x_1-\hat{x}_1)^2 + (x_2-\hat{x}_2)^2]$$

$$= \frac{1}{2}\,E\left[\left(\frac{1}{f'(x_1)}\right)^2 \delta_1^2 + \left(\frac{1}{f'(x_2)}\right)^2 \delta_2^2\right]$$

$$\le \frac{1}{2}\left\{ E\left[\left(\frac{1}{f'(x_1)}\right)^2 16\epsilon^2 + .354\right] \Pr[\hat{r}=r] \right.$$

$$\left. + \ E\left[\left(\frac{1}{f'(x_1)}\right)^2 (4) + \left(\frac{1}{f'(x_2)}\right)^2 (3)\right] \Pr[\hat{r}\neq r] \right\}$$

---

*See end of this appendix for bound on

$$\left|\frac{g''(\tau_i)}{2!}\right|$$

Let $P_e \triangleq \Pr[\hat{r} \neq r]$ (probability of error). Also, since

$$f(x) = \frac{2}{\pi} \tan^{-1} x = \theta ,$$

then

$$E\left(\frac{1}{f'(x)}\right)^2 = \frac{3}{2} \pi^2 \quad .$$

Thus

$$J_{NL}(\sigma^2) \leq B(\sigma^2) \quad ,$$

where

$$B(\sigma^2) = \frac{1}{2} \left\{ \left[\frac{3}{2} \pi^2 16\sigma^2 + .354 \right] (1-P_e) + \left[10 \frac{1}{2} \pi^2 \right] P_e \right\} \quad .$$

Bound on $|g''(\tau)|$: Let $\tau \in [\theta, \hat{\theta}]$ and $\delta = \hat{\theta} - \theta$.

$$g(\theta) = \tan \frac{\pi}{2} \theta$$

$$g'(\theta) = \frac{\pi}{2} \left(1 + \tan^2 \frac{\pi}{2} \theta\right)$$

$$g''(\theta) = \pi \tan \frac{\pi}{2} \theta \; g'(\theta)$$

$$= \frac{\pi^2}{2} \tan \frac{\pi}{2} \theta \left(1 + \tan^2 \frac{\pi}{2} \theta\right)$$

$$= \frac{\pi^2}{2} \tan \frac{\pi}{2} \theta \left(\sec^2 \frac{\pi}{2} \theta\right)$$

or

$$g''(\theta) = \pi \, g(\theta) \, g'(\theta) \quad . \qquad \text{(3C.5)}$$

Thus, we see from Figures 3C. 3 and 3C. 4 that

$$|g(\tau)| \leq \max\{|g(\theta)|, |g(\hat{\theta})|\} \triangleq A(\theta, \hat{\theta})$$

$$|g'(\tau)| \leq \max\{|g'(\theta)|, |g'(\hat{\theta})|\} \triangleq B(\theta, \hat{\theta}) \quad .$$

From (3C. 5),

$$|g''(\tau)| \leq \pi A(\theta, \hat{\theta})\, B(\theta, \hat{\theta}) \quad .$$

Therefore, $|g''(\tau)|$ will not blow up as δ gets smaller, since A and B do not blow up as δ gets smaller.

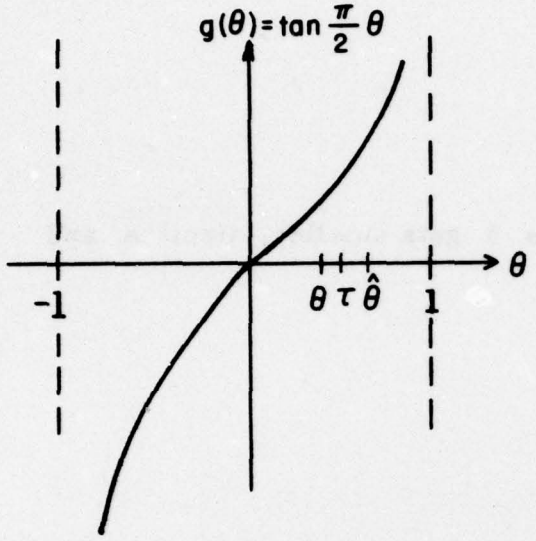


FIG. 3.C.1 BOUND ON $|g(\tau)|$

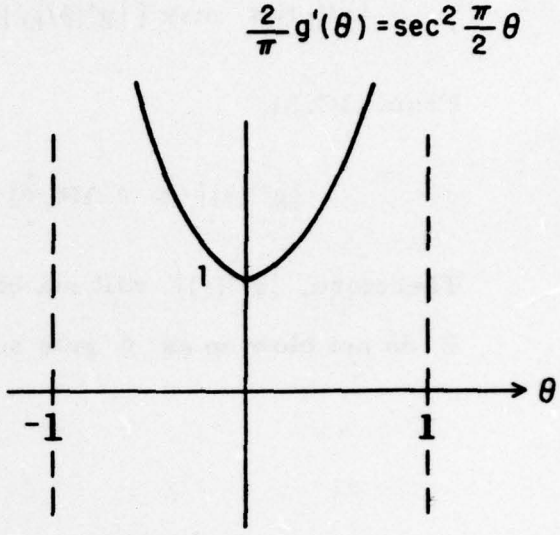


FIG. 3.C.2 BOUND ON $|g'(\tau)|$

CHAPTER IV

CONCLUSION

The main features of the Spence problem of Chapter II, from a decision and control point of view, are the dynamic information structure (i.e., signaling) and the multiple equilibria. Multiple Nash equilibria in a noncooperative NZS game are undesirable, because DM1 may choose one equilibrium strategy, say γ_1^a, but DM2 might not choose the corresponding γ_2^a, but some other equilibrium γ_2^b. The pair (γ_1^a, γ_2^b), in general, is not an equilibrium. By assuming certain parameters were fixed instead of variable, we avoided this problem and obtained a unique equilibrium.

As a vehicle for insights, the model was set up as a two-person decision problem. This allowed us not only to find new solutions, but also to handle modifications of the problem more easily. For example, we defined an adjustment procedure for each decision maker and proved sufficient conditions for stability. We also investigated threshold effects and found that, under certain circumstances, signaling ceases when different parameters in the problem are varied. The main results were that if signaling cost or signaling noise are too high, or if the variability of the underlying unknown signal is too low, then signaling is not worthwhile. Therefore, from what originally appeared as a very simple example, a tremendous richness of detail and insight have emerged.

Extending the decision theory framework, we saw in Chapter III how Shannon information theory can be modeled as a two-person team problem with signaling. This set-up allowed us to discuss coding problems in which the delay between emission of source symbols and transmission of coded signals was fixed. This was called "real-time information theory," suggesting applications where coding precedes actions which must be taken within a specified time. Since Shannon's theorem states the best we can do with no delay restriction, it provides a bound against which we can judge the performance of the real-time scheme. General results about performance were also derived; for example, repetition of source symbols as an encoding scheme is as good as the best linear encoder. However, if the block lengths of the source and channel are equal, then for both variable and fixed delay, linear encoders and decoders are optimal; that is, they attain the Shannon bound. If the block lengths are not equal, then for fixed delay, linear may not be optimal.

The major contribution of this work is not to prove significantly new results, but rather to unify the disparate fields of team theory, market signaling in economics, information structures, and classical information theory. Hopefully, the general conceptual framework presented here will encourage joint efforts among researchers in these separate fields.

ACKNOWLEDGMENTS

The author would like to thank Prof. Y. C. Ho, Prof. E. Wong, and Prof. M. Spence for their helpful suggestions and important contributions to this report.

JSEP REPORTS DISTRIBUTION LIST

N00014-75-C-0648

Defense ... mission Center
Attn: DDC-TCA (Mrs. V. Caponio)
Cameron Station
Alexandria, Virginia 22314

Asst. Director
Electronics and Computer Sciences
Office of Director of Defense Research and Engineering
The Pentagon
Washington, DC 20315

Office of the Director of Defense Research and Engineering
Information Office Library Branch
The Pentagon
Washington, DC 20301

ODDR and E Advisory Group on Electron Devices
201 Varick Street
New York, New York 10014

Chief, R and D Division (360)
Defense Communications Agency
Washington, DC 20301

Director, *National Security Agency*
Fort George G. Meade, Maryland 20755
Attn: Dr. T. J. Beahn

Institute for Defense Analysis
Science and Technology Division
400 Army-Navy Drive
Arlington, Virginia 22202

Dr. Stickley
Defense Advanced Research Projects Agency
Attn: Technical Library
1400 Wilson Boulevard
Arlington, Virginia 22209

Dr. R. Reynolds
Defense Advanced Research Projects Agency
Attn: Technical Library
1400 Wilson Boulevard
Arlington, Virginia 22209

AF/RDPS
The Pentagon
Washington, DC 20330

AFSC (LJ/Mr. Irving R. Mirman)
Andrews AFB
Washington, DC 20334

Director of Electronics and Weapons
HQ AFSC/DLC
Andrews AFB, Maryland 20334

Directorate of Science
HQ AFSC/DLS
Andrews AFB
Washington, *DC 20334*

LTC J. W. Gregory
AF Member, TAC
Air Force Office of Scientific Research
Bolling AFB
Washington, DC 20332

Mr. Carl Sletten
AFCRL/LZ
Hanscom AFB, MA 01731

Dr. Richard Picard
AFCRL/OPL
Hanscom AFB, MA 01731

Mr. Robert Barrett
AFCRL/LQ
Hanscom AFB, MA 01731

Dr. John N. Howard
AFCRL (CA)
Hanscom AFB, MA 01731

Dr. Richard B. Mack
AFCRL/LZR
Hanscom AFB, MA 01731

Documents Library (TILD)
Rome Air Development Center
Griffiss AFB, New York 13441

Mr. H. E. Webb, Jr. (IBCP)
Rome Air Development Center
Griffiss AFB, New York 13441

Mr. Murray Kesselman (IBCA)
Rome Air Development Center
Griffiss AFB, New York 13441

Mr. W. Edwards
AFAL/TE
Wright-Patterson AFB, Ohio 45433

Mr. R. D. Larson
AFAL/DHR
Wright-Patterson AFB, Ohio 45433

Chief Scientist
AFAL/CA
Wright-Patterson AFB, Ohio 45433

HQ ESD (DRI/Stop 22)
Hanscom AFB, MA 01731

Professor R. E. Fontana
Head Dept. of Electrical Eng.
AFIT/ENE
Wright-Patterson AFB, Ohio 45433

Mr. John Mottsmith (MCIT)
H... ESD (AFSC)
Hanscom AFB, MA 01731

LTC Richard J. Gowen
Professor
Dept. of Electrical Engineering
USAF Academy, Colorado 80840

AUL/LSE-9661
Maxwell AFB, Alabama 36112

AFETR Technical Library
P.O. Box 6698, MU 5650
Patrick AFB, Florida 32542

ADTC (DLOSL)
Eglin AFB, Florida 32542

HQ AMD (RDR/COL Godden)
Brooks AFB, Texas 78235

USAF European Office of Aerospace Research
Technical Information Office
Box 14, FPO, New York 09510

Dr. Carl E. Baum
AFWL (DE)
Kirtland AFB, New Mexico 87117

ASAFSAM/RAL
Brooks AFB, Texas

HQDA (DAMA-ARZ-A)
Washington, DC 20310

Commander
US Army Security Agency
Attn: IARD-T
Arlington Hall Station
Arlington, VA 22212

Commander
US Army Materiel Development and Readiness Command
Attn: Technical Library, Rm 7S 35
5001 Eisenhower Avenue
Alexandria, VA 22333

Commander
US Army Ballistics Research Laboratory
Attn: DRXBR-BAD
Aberdeen Proving Ground
Aberdeen, MD 21005

Commander
Picatinny Arsenal
Attn: SMUPA-TS-T-S
Dover, NJ 07801

US Army Research Office
Attn: Dr. Hermann Robl
P.O. Box 12211
Research Triangle Park, NC 27709

US Army Research Office
Attn: Mr. Richard O. Ulsh
P.O. Box 12211
Research Triangle Park, NC 27709

US Army Research Office
Attn: Dr. Jimmie R. Suttle
P.O. Box 12211
Research Triangle Park, NC 27709

US Army Research Office
Attn: Dr. Horst Wittmann
P.O. Box 12211
Research Triangle Park, NC 27709

Commander
Frankford Arsenal
Attn: Mr. George, C. White, Jr.
Deputy Director, Pitman-Dunn Lab.
Philadelphia, PA 19137

Commander
US Army Missile Command
Attn: Chief, Document Section
Redstone Arsenal, AL 35809

Commander
US Army Missile Command
Attn: DRSMI-RR
Redstone Arsenal, AL 35809

Commander
US Army Materials and Mechanics Research Center
Attn: Chief, Materials Sciences Division
Watertown, MA 02172

Commander
Harry Diamond Laboratories
Attn: Mr. John E. Rosenberg
2800 Powder Mill Road
Adelphi, MD 20783

Commandant
US Army Air Defense School
Attn: ATSAD-T-CSM
Fort Bliss, TX 79916

Commandant
US Army Command and General Staff College
Attn: Acquisitions, Lib. Div.
Fort Leavenworth, Kansas 66027

Dr. Hans K. Ziegler
Army Member, TAC/JSEP
US Army Electronics Command
(DRSEL-TL-D)
Fort Monmouth, NJ 07703

Mr. J. E. Teti
Executive Secretary, TAC/JSEP
US Army Electronics Command
(DRSEL-TL-DT)
Fort Monmouth, NJ 07703

Director
Night Vision Laboratory, ECOM
Attn: DRSEL-NV-D
Fort Belvoir, VA 22060

Commander/Director
Atmospheric Sciences Laboratory (ECOM)
Attn: DRSEL-BL-DD
White Sands Missile Range, NM 88002

Director
Electronic Warfare Laboratory (ECOM)
Attn: DRSEL-WL-MY
White Sands Missile Range, NM 88002

Commander
US Army Armament Command
Attn: DRSAR-RD
Rock Island, IL 61201

Director, Div. of Neuropsychiatry
Walter Reed Army Institute of Research
Washington, DC 20012

Commander
USASATCOM
Fort Monmouth, NJ 07703

Commander
US Army R and D Group (Far East)
APO San Francisco, CA 96343

Commander
US Army Communications Command
Attn: Director, Advanced Concepts Office
Fort Huachuca, AZ 85613

Project Manager
ARTADS
EAI Building
West Long Branch, NJ 07764

Commander
US Army White Sands Missile Range
Attn: STEWS-ID-R
White Sands Missile Range, NM 88002

Director, TRI-TAC
Attn: TT-AD (Mrs. Briller)
Fort Monmouth, NJ 07703

Commander
US Army Communication Command
Attn: CC-OPS-PD
Fort Huachuca, AZ 85613

COL Robert Nuce
Senior Standardization Representative
US Army Standardization Group, Canada
Canadian Force Headquarters
Ottawa, Ontario, Canada *K1A OK2*

Commander
US Army Electronics Command
Attn: DRSEL-RD-O (Dr. W. S. McAfee)
CT-L (Dr. H. Bauer)
NL-O (Dr. H. S. Bennett)
NL-T (Mr. R. Kulinyi)
TL-B
VL-D
WL-D
TL-MM (Mr. N. Lipeta)
NL-H (Dr. F. Schwering)
TL-E (Dr. S. Kronenberg)
TL-E (Dr. J. Kohn)
TL-I (Dr. C. Thornton)
NL-B (Dr. S. Amoroso)
Fort Monmouth, NJ 07703

Project Manager
Ballistic Missile Defense Program Office
Attn: DACS-BMP (Mr. A. Gold)
1300 Wilson Blvd.
Washington, DC 22209

Dr. Sidney Ross
Technical Director
SARFA-TD
Frankford Arsenal
Philadelphia, PA 19137

Department of Naval Research
Electronic and Solid State Sciences Program (Code 427)
800 N. Quincy
Arlington, Virginia 22217

Office of Naval Research
Code 200
Assistant Chief for Technology
800 N. Quincy
Arlington, Virginia 22217

Office of Naval Research
Information Sciences Program
(Code 437)
800 N. Quincy
Arlington, Virginia 22217

Naval Research Laboratory
4555 Overlook Avenue, S.W.
Washington, DC 20375
Attn: Codes 2627
4000
4105
5000
5200
5203
5210
5270
5300
5400
5460
5464
5500
5510
6400

Director
Office of Naval Research Branch Office
536 South Clark Street
Chicago, Illinois 60605

San Francisco Area Office
Office of Naval Research
760 Market Street, Room 447
San Francisco, CA 94102

Dr. A. Laufer
Chief Scientist
Office of Naval Research Branch Office
1030 East Green Street
Pasadena, CA 91101

Director
Office of Naval Research Branch Office
715 Broadway, 5th Floor
New York, New York 10003

Mr. L. W. Sumner
Naval Electronics Systems Command
NC No. 1
2511 Jefferson Davis Highway
Arlington, VA 20360

Mr. R. Freslin
Naval Electronics Systems Command
NC No. 1
2511 Jefferson Davis Highway
Arlington, VA 20360

Mr. N. Butler
Naval Electronics Systems Command
NC No. 1
2511 Jefferson Davis Highway
Arlington, VA 20360

Dr. H. J. Mueller
Naval Air Systems Command
JP No. 1
1411 Jefferson Davis Highway
Arlington, VA 20360

Capt. R. B. Meeks
Naval Sea Systems Command
NC No. 3
2531 Jefferson Davis Highway
Arlington, VA 20362

Commander
Naval Surface Weapons Center
Attn: Technical Library
Silver Spring, Maryland 29010

Officer-in-Charge
Naval Surface Weapons Center
Dahlgren Laboratory
Dahlgren, Virginia 22448

Naval Air Development Center
Attn: Technical Library
Johnsville
Warminster, Penn. 18974

Commander
Naval Avionics Facility
Indianapolis, Indiana 46241
Attn: D/035 Technical Library

Naval Missile Center
Technical Library
Code 5632.2
Point Mugu, CA 93042

Naval Weapons Center
Attn: Technical Library
Code 533
China Lake, CA 93555

Naval Weapons Center
Attn: Code 6010
China Lake, CA 93555

Naval Weapons Center
Mathematics Division
China Lake, CA 93555

Naval Training Equipment Center
Technical Library
Orlando, Florida 32813

Naval Research Laboratory
Underwater Sound Reference Div.
Technical Library
P.O. Box 8337
Orlando, Florida 32806

Naval Underwater Sound Center
Technical Library
New London, Conn. 06320

U.S. Naval Oceanographic Office
Library - Code 1600
Washington, DC 20373

Commandant, Marine Corps
Scientific Advisor (Code AX)
Washington, DC 20360

Dr. Gernot M. R. Winkler
Director, Time Service
US Naval Observatory
Mass. Ave. at 34th St. N.W.
Washington, DC 20390

Naval Postgraduate School
Technical Library
Monterey, California 93940

Naval Electronics Laboratory Center
Technical Library
San Diego, California 92152

Naval Electronics Laboratory Center
Attn: Code 2600
San Diego, CA 92152

Naval Electronics Laboratory Center
San Diego, California 92152
Attn: Codes 2000
2200
2300
3000
3200
3300
4800
5000
5200
5300

Naval Undersea Center
Technical Library
San Diego, California 92152

Naval Ship Research and Development Center
David W. Taylor
Code 522.1
Bethesda, Maryland 20084

Office of Chief of Naval Operations
NAICOM/MIS Planning Branch
NOP-916D, Pentagon
Washington, DC 20350

Mr. F. C. Schwenk, RD-T
National Aeronautics and Space Administration
Washington, DC 20546

Los Alamos Scientific Laboratory
Attn: Reports Library
P.O. Box 1663
Los Alamos, New Mexico 87544

M. Zane Thornton
Deputy Director Institute for Computer Sciences and Technology
National Bureau of Standards
Washington, DC 20234

Director, Office of Postal Technology (R and D)
US Postal Service
11711 Parklawn Drive
Rockville, Maryland 20852

NASA Lewis Research Center
Attn: Library
21000 Brookpark Road
Cleveland, Ohio 44135

Library -851
Bureau of Standards Acquisition
Boulder, Colorado 80302

MIT Lincoln Laboratory
Attn: Library A-082
P.O. Box 73
Lexington, Mass. 02173

Dr. Jay Harris
Program Director
Devices and Waves Program MEF
1800 G Street
Washington, DC 20550

Dr. Howard W. Etzel
Deputy Director
Div. of Materials Research MEF
1800 G Street
Washington, DC 20550

Dr. Dean Mitchell
Program Director
Solid-State Physics
Div. of Materials Research
National Science Foundation
1800 G Street
Washington, DC 20550

Director
Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Director
Microwave Research Institute
Polytechnic Institute of New York
Long Island Graduate Center, Route 110
Farmingdale, New York 11735

Asst. Director
Microwave Research Institute
Polytechnic Institute of New York
333 Jay Street
Brooklyn, New York 11201

Director
Columbia Radiation Laboratory
Dept. of Physics
Columbia University
538 West 120th Street
New York, New York 10027

Director
Coordinated Science Laboratory
University of Illinois
Urbana, Illinois 61801

Director
Stanford Electronics Laboratory
Stanford University
Stanford, California 94305

Director
Microwave Laboratory
Stanford University
Stanford, California 94305

Director
Electronics Research Laboratory
University of California
Berkeley, California 94720

Director
Electronics Sciences Laboratory
University of Southern California
Los Angeles, California 90007

Director
Electronics Research Center
The University of Texas at Austin
Engineering-Sciences Bldg. 112
Austin, Texas 87812

Director of Laboratories
Div. of Engineering and Appl. Physics
Harvard University
Pierce Hall
Cambridge, Mass. 02138