A CASE STUDY IN THE ECONOMICS OF INFORMATION AND COORDINATION

THE WEATHER FORECASTING SYSTEM

Richard R. Nelson

Sidney G. Winter, Jr.

September 1963

P-2788

A CASE STUDY IN THE ECONOMICS OF INFORMATION AND COORDINATION THE WEATHER FORECASTING SYSTEM

Richard R. Nelson*

The RAND Corporation, Santa Monica, California

and

Sidney G. Winter, Jr.*

University of California, Berkeley; Consultant to The RAND Corporation

The economics of information and coordination has long occupied a paradoxical position in the structure of economic thought as a whole. In both the classical and modern discussion of the allocative efficiency of a system of prices and markets, great emphasis has been put on the economy of information, communication and coordination that such a system may achieve. Yet, with few exceptions, the rigorous formulations of the theory of such a system do not contain reference to any real costs of information processing. Business firms and consumers are assumed to _have_ all relevant information. The computations that enable firms to make optimum choices of technique and output are assumed to require no investment of time or resources; the process of maximizing utility does not itself involve disutility; the markets where supply and demand are balanced function without benefit of land and buildings, clerical assistance, telephones, and so forth; the price changes generated by

market adjustment are costlessly known by all and summarize all of the information firms and consumers need to know to make their decisions. Thus the analysis of the ability of a price and market system to achieve economies of information processing typically has been carried on in a context of assumptions which makes information a free good, the processing of it costless, and the attempt to economize it pointless.

The paradox is rendered even more striking when it is realized that activities involving the production and distribution of information (education and basic research, for example) provide some of the most widely recognized examples of external economies and indivisibilities, phenomena which cause serious difficulties in the achievement of efficiency in a price and market system.

It is clear that for many reasons it is important to try to treat the economics of information and coordination more carefully and systematically than it generally is. It must be recognized that "prices" are not the only information of economic significance. The costs of generating, sending, and processing information must be considered. There are some special problems relating to information as a commodity which require special treatment.

Some beginnings on the construction of a framework capable of dealing with certain aspects of information and coordination have been made in recent years on foundations contributed largely by statistical decision theory. Of particular importance is the work of J. Marschak,[1]

---

[1] J. Marschak, "Towards an Economic Theory of Organization and Information," in Decision Processes, Thrall, Davis and Coombs, eds., (New York: John Wiley & Sons, 1954); "Elements for a Theory of Teams," Management Science, Vol. 1 (January 1955), pp. 127-37; "Theory of an Efficient Several-Person Firm," American Economic Review, Vol. 50 (May 1960), pp. 541-548.

R. Radner,[2] C. B. McGuire[3] and others on the economic theory of "teams" -- a "team" being defined as an organization composed of individual decision makers all of whom share a single common objective. T. Marschak has examined the effects of slowness of information processing on economic organization.[4] G. Stigler[5] and R. Nelson[6] have explored other aspects of the economics of information. Other inquiries into the economics of information have been stimulated by the increasing role of research and development activities in the economy, and the resulting interest in questions of how such activities are and should be carried on.

But very little has been done in the way of studies of particular applied problems in the economics of information and coordination to test the usefulness of the theory and to indicate areas where it needs strengthening. In the present paper, we adapt the concepts of the emerging economic theory of information to examine some particular

---

[2] R. Radner, "The Linear Team: An Example of Linear Programming Under Uncertainty," Proc. Second Symposium on Linear Programming, (Washington: National Bureau of Standards, 1955); "The Application of Linear Programming to Team Decision Problems," Management Science, Vol. 5, pp. 143-150; "The Evaluation of Information in Organizations," Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability (University of California Press, 1961), pp. 491-533; "Team Decision Problems," The Annals of Mathematical Statistics, Vol. 33 (September 1962), pp. 857-881.

[3] C. B. McGuire, "Some Team Models of a Sales Organization," Management Science, Vol. 7 (January 1961), pp. 101-130. Also, "Comparisons of Information Structures," Cowles Foundation Discussion Paper No. 71, April 1959 (unpublished).

[4] T. Marschak, "Centralization and Decentralization in Economic Organizations," Econometrica, Vol. 27 (July 1959), pp. 399-430.

[5] G. Stigler, "The Economics of Information," Journal of Political Economy, June 1961.

[6] R. Nelson, "Uncertainty, Prediction and Competitive Equilibrium," Quarterly Journal of Economics, February 1961, pp. 41-62.

problems involving the dissemination, use, and evaluation of weather information. The theoretical framework follows closely that of Marschak and Radner in the works referred to earlier but incorporates several new elements which we found necessary to deal with our problem. The example we use is more than an illustration, but less than a full scale case study. While it originally was designed as part of a study examining some aspects of the weather forecasting system per se,[7] this case is presented here in the belief that examination of a real problem should shed some light on the economics of information generally. In particular, we believe it raises some important questions which hitherto have gone unnoticed.

I. THE GENERAL FRAMEWORK

In trying to deal with the economics of weather forecasting, we were forced to modify in some respects the framework developed in connection with the theory of teams, but basically found that framework a useful one to work within.

Looking first at the user of information, we consider a single decision maker whose problem is to make a choice of an action a from some set A of possible actions. From this choice the decision maker will derive a utility u which depends in some known way upon the "true state of the world," w, where w is an element of a set W of possible true states of the world. That is

$$u = \gamma(a, w) \qquad a \in A,\ w \in W$$

---

[7] We would like to acknowledge our indebtedness to R. Rapp, D. Sartor, and L. Kolb for their cooperation in the study which produced these "case studies." The original study from which the case is drawn is our "Weather Information and Economic Decisions: A Preliminary Report," RAND RM-2620-NASA, August 1960.

In the case where A and W are finite sets, the function ν can of course be displayed as a "payoff matrix" of the type made familiar by statistical decision theory and the theory of games.

The decision maker does not in general know the true state of the world, w. But an information source supplies him with some "message" y, where y is an element of a set Y of possible messages.[8] There exists some joint probability distribution $\pi(y, w)$ over the sets Y and W of messages and true states of the world, and we assume that this distribution is known to the decision maker. We further assume that the decision maker will, upon receiving message y, choose that action a which maximizes expected utility for the conditional probability distribution of w given y.[9] That is, he will choose a to maximize.

$$E(u|y) = E_{w|y} \; \nu(a, w)$$

The function $\hat{\alpha}$ which associates with each possible message y the action a which maximizes $E(u|y)$ is the decision maker's best (Bayesian) decision rule, or strategy. Thus

$$a = \hat{\alpha}(y)$$

[8]We will assume that the set Y is defined in such a way that the message y is the only information the decision maker receives. Thus the "information source" may consist in part of the decision maker's own activities. In the formal framework this is merely a convention, but in our case studies we will turn it into a genuine assumption by ignoring any information the decision maker has that is not supplied by the weather forecaster.

[9]Of course, on the modern view of utility, the question of the validity of this behavioral assumption is more properly a question about the existence of the utility function.

Given the conditional probability distribution of w for each y, and the marginal probability distribution of y, and assuming the optimum decision rule is used, overall expected utility is:

$$E(\hat{u}) = E_y E_{w|y} V[\hat{\alpha}(y), y]$$

Turning next to analysis of the information source; this source may be managed by another member of the same team to which the decision maker belongs.[10] Or it may not; in the case of weather forecasting the user of the forecasts, and the forecaster, need not have identical utility functions. However we are interested in decisions of the information source which are optimal from the point of view of the information user, taking proper account of the costs of the information source and assume that institutional arrangements can be worked out so that decisions of this sort, if they can be identified, will in fact be made. Of course, one of the tasks of the theory of information and coordination is to discover the institutional arrangements which meet this objective.

We regard the activities of the information source as involving two stages. First there is an <u>observation</u> stage, in which some element z of a set Z of possible observations is observed. There is assumed to be a joint probability distribution, $\phi(z, w)$ over the sets Z and W of possible observations and true states of the world. This joint distribution is not necessarily known to the decision maker, but it is known to the manager of the information source. Secondly, there is a

[10] More precisely, the case of a two member team in which one member does all the "observing" and the other does all the "acting."

communication stage, in which the message y is determined as a function of the observation z, $y = \eta(z)$, and the message is made available to the decision maker. Following J. Marschak, we call the function η the information structure.[11]

The probability distribution ϕ and the information structure η obviously determine between them the joint probability distribution π of messages and true states of the world. Both of these functions are to some extent subject to choice by the manager of the information source.[12] The distribution ϕ is determined by decisions relating to what shall be observed and how these things shall be observed, as well as by the distribution of true states of the world.[13] As for the

[11]Actually, in the typical team theory formulation, the information structure is a mapping from the set of states of the world to the set of messages. At the purely formal level, it is not necessary to hypothesize a set of possible observations distinct from the set of states of the world, since such problems as errors in observations can always be handled by supposing the values of error terms to be included in a complete description of the true state of the world. However, it is often convenient to hypothesize a separate observation stage, for the distinction between observation and transmission of information corresponds to an obvious distinction between the types of equipment and skills involved in different stages of the total information acquiring-and-transmitting process. We have found this split convenient in our study.

[12]Both our ϕ and our η would be regarded as determinants of the information structure in the usual team theory formulation, and the choice of an information structure would therefore comprise choices with respect to both functions.

[13]To take an example which illustrates the connection between these concepts and ordinary statistical decision problems, let w be a parameter of some probability distribution of known form. Let the observation z be a statistic computed from a sample drawn from a population with this distribution, but with the parameter w shifting from one time period to the next. Then the joint distribution ϕ is obviously influenced by the statistic chosen and by the sampling method and sample size.

information structure, there may exist possibilities of communicating z itself, of communicating only the sign of z, or of partitioning the range of values of z into intervals and communicating a signal which tells the interval in which z falls. In the examples to be considered subsequently, our primary focus will be on the problem of choosing η, with ϕ being taken as given.

In many instances, it will be a satisfactory approximation to assume that utility is linear in money, since the amount the decision maker has at risk in any particular confrontation with his decision problem may be small relative to his total capital. We then lack only a determination of a zero point to permit us to define the value of a particular information structure (with ϕ taken as given) to the decision maker. We follow Marschak and Radner in defining

$$V(\eta) = E(\hat{u}) - \underset{a}{\text{Max}}\, E_w\, \nu(a, w)$$

$$= E_y\, E_{w|y}\, \nu(\hat{\alpha}(y), y) - E_w\, \nu(\hat{a}, w)$$

That is, we assign the value 0 to the information structure which provides no information at all to the decision maker, assuming that in this case the decision maker knows the probability distribution of true states of the world (a special case of his knowing the distribution π) and that he always takes the action which maximizes expected payoff relative to that probability distribution. The quantity $V(\eta)$ is the value of the information structure to the decision maker in the sense of demand price, when the only alternative available is no information at all. Obviously, the other alternatives available and the costs of all of the alternative information structures are also determinants of

the value of an information structure in the sense of the demand price of that particular information service.

The null information structure that provides no information at all provides one point of comparison for alternative information structures; another is provided by the "identity" information structure that provides the observation z itself: $y = \eta(z) = z$. Denoting this information structure by η^*, we have

$$V(\eta^*) = E_z E_{w|z} \upsilon(\hat{\alpha}(z), w)$$

If an information structure η has the same value as η^*, we will refer to η as a <u>sufficient</u> information structure -- each of the signals y yielded by such an information structure must be a "sufficient statistic," i.e., it must convey all the information bearing on the decision problem that the observation z contains.

In a situation where the cost of the information system is independent of η, obviously η^* or any other sufficient information structure is optimal. In a situation where there are costs involved in sending information, it may not be optimal to send "sufficient" information. While it is extremely difficult to get at all the factors which determine the costs of an information structure, one facet which might be considered is the number of possible messages which η involves.

In the case where the set A of possible actions contains only a finite number n of elements, it is easily seen that there exists a sufficient information structure such that no more than n messages have

positive probability.[14] For we can simply let the message $\eta(z)$ be y_i whenever z is an observation such that $\hat{\alpha}(z)$ is a_i; then clearly $\hat{\alpha}(y_i)$ is a_i also, and the decision maker responds to each of the messages $y_1, \ldots, y_n$ just as he would if he had full information about z. To put it another way, if n different messages can be given to the decision maker, there exists an information structure such that he will never take an action which is inferior to the action he would take given complete knowledge of z.[15]

To the extent that the cost of η is determined by the number of possible messages, it is clear that (abstracting from problems of noise) there is no reason to choose an information structure which involves more than n possible messages. Of course it may be optimal to use an η which involves less than n possible messages. In such cases it is important to note that if we look at information structures involving n-k messages, there exists at least one such that the decision maker will never take an action which is worse than the $k + 1^{st}$ best action when actions are ranked according to their expected payoff given full

[14]Formally, all possible information structures have the same codomain, the set Y. When we refer subsequently to the number of messages "associated" with an information structure, we mean the number of messages in the range of that information structure; i.e., the number of messages y for which there exists a possible observation z such that $\eta(z) = y$. A further complication is that for some distributions ϕ the relevant z or z's may never occur, so that not all messages in the range of η need have positive probability. These complications should not give rise to any confusion.

[15]It should be noted that one message may be conveyed without any physical signal; i.e., the absence of a signal (telephone call or whatever) may have a standard interpretation as a message. See the discussion on p. 27 below.

knowledge of z. Furthermore, among the information structures yielding this value to the decision maker, there is at least one in which only n-k messages have positive probability. These propositions are the starting point for solving the problem of determining the information structure which will maximize the decision maker's expected payoff when the gains to the decision maker must be balanced off against a cost which depends only on the number of different messages provided or when there is a constraint on the number of different messages that can be communicated.

There are several different types of applications for the foregoing analysis. It may serve as the basis of a positive theory of behavior under uncertainty and of the value of information. Or it may be regarded as a normative theory, pointing the way toward the determination of best decision rules and best information structures in particular situations. Or the theory of the behavior of the decision maker may be regarded as descriptive, while the remainder of the analysis is regarded as normative guidance for the manager of the information source. It is the latter point of view that we take in our subsequent discussion of the principles that should guide a "public forecaster" (e.g., the Weather Bureau).

II. THE USE OF WEATHER INFORMATION[16]

We turn now to an application of the general framework in which we identify w, the true state of the world, with some future weather condition, the observation z with the meteorological information available to a weather forecaster, and the message y with a weather forecast of some sort. (This last identification involves an assumption that the weather forecast is the decision maker's sole source of weather information.) The assumption that the decision maker knows the probability distribution of true states of the world, even if he has no other information, translates into an assumption that the decision maker has accurate climatological information for those weather events which are relevant to his problem; i.e., he knows the relative frequencies with which these events will occur given the time of the

---

[16] We should emphasize at this point that we are by no means the first to apply some sort of statistical decision theory framework to the problems of making, responding to and evaluating weather forecasts. Among the prior contributions in this field we may cite, I. Gringorten, "Probability Estimates of the Weather in Relation to Operational Decisions," Journal of Meteorology, Vol. 16 (December 1959), pp. 663-71; and "On the Comparison of One or More Sets of Probability Forecasts," Journal of Meteorology, Vol. 15 (June 1958), pp. 283-87. Also, J. C. Thompson and G. L. Brier, "The Economic Utility of Weather Forecasts," Monthly Weather Review, Vol. 83 (November 1955), pp. 249-53; J. C. Thompson, "On the Operational Deficiencies in Categorical Weather Forecasts," Bulletin of the American Meteorological Society, Vol. 33 (June 1952), pp. 223-26. The latter papers consider in particular the simple "protect - don't protect" problem which serves as our main example. Our own treatment casts this example in a more general framework and also provides a particular illustration of it. The present paper draws in part on our own earlier work on the subject, reported in "Weather Information and Economic Decisions: A Preliminary Report," The RAND Corporation, RM-2620-NASA, August 1, 1960.

year, but, in the absence of weather forecasts, cannot predict better than this.

Our example is representative of a wide class of real world decision problems which can, without too much violence to reality, be reduced to the following simple structure: the decision maker must choose between taking or not taking some specific protective action against a future unfavorable weather condition; taking the protective action involves some cost with certainty; not taking it involves escaping that cost but incurring a certain loss if the unfavorable weather condition does in fact occur.

Thus a newspaper distributor, who has a standard routine for distribution, can wrap his papers in wax paper to protect them from rain. A storekeeper can tape his windows to protect them from a threatening hurricane. A citrus grower can light smudge pots to protect his fruit from frost.

We will examine here a decision problem that confronts the dispatcher of a fleet of trucks.

The schedule requires that all loading be accomplished on the day or evening before dispatch so that the trucks can leave early in the morning. All of the trucks are "flat racks" (uncovered trucks), and the merchandise carried is seldom of a sort which will be damaged significantly by light rain. It is customary to ship such merchandise -- building material, canned goods, etc. -- unprotected or protected only by normal packaging, if the chances of rain are judged to be slim.

However, a moderate or heavy rain can cause considerable damage even to cargo usually considered quite unperishable. The top layers of

the goods may be completely soaked. Cardboard packaging may be weakened to the point of being worthless. Labels, for example, on canned goods, may be loosened and even soaked off. A reasonable figure for the average damage resulting from a moderate or heavy rain is $500 per truckload of unprotected merchandise.

If a significant amount of rain is expected, the dispatcher can direct his crews to "tarp the trucks" (cover the merchandise with a tarpaulin) after loading them. The cost of doing this job, including labor time plus wear and tear on the tarping on an average trip, may be put at about $20 a truck. To permit an early morning departure, tarping must be done the prior evening.

For the purposes of this study we consider any amount of rain in excess of .15 inches as "heavy," and assume that if rain is less than this amount no damage will result. Each evening the dispatcher must decide whether or not to tarp the trucks. Clearly it would be good policy to tarp the trucks only on evenings preceding bad days and at no other times, but in the absence of perfect forecasts this is impossible. In the absence of perfect information, the dispatcher bases his decision on the weather forecast he receives from a private weather consultant. If the forecast is for rain, the trucks are tarped, but otherwise not.

In the Los Angeles area, the site of our example, the climatological probability of rain in excess of .15 inches is .09 during the winter rainy season, the period of our example. It is obvious that if the dispatcher has to make his daily decisions on this information alone and aims to minimize expected cost, he should order the trucks tarped every evening, thus incurring a daily cost of $20 per truck. The

alternative, never tarping the trucks, would work out well on an average of 91 days out of 100, but on an average of nine days out of 100 the company would incur a loss of $500 per truck, an average daily loss of $45 per truck. (We assume, not perfectly accurately, that regardless of how the weather looks on the morning of departure, no makeshift protective measures can be taken and the trucks must depart on schedule.)

Let us now consider the forecasts the decision maker receives from a private weather consultant. In our trucking example the forecast received by the decision maker is either a prediction that weather state w_1 = "heavy rain" will occur -- call this forecast y_1 -- or a forecast that weather state w_2 (= less than .15 inches rain) will occur -- call this forecast y_2. Records show that out of 100 forecasts made during a four-month period, 18 were predictions of rain, and 82 were predictions that it would not rain. The no-rain forecasts were very reliable; on only two days did rain occur when none was forecast. This high accuracy of the y_2 forecasts was achieved because the forecaster always predicted rain whenever there was any doubt in his mind. His y_1 forecasts, therefore, were not very reliable. Indeed, on 11 out of the 18 days that he forecast rain, it did not rain. It will be demonstrated later that this low accuracy of the y_1 forecasts, rather than reflecting adversely on the forecaster, is an indication that he was sensibly tailoring his forecasts to his customer's needs.

The forecasts thus define the following two-by-two contingency table:

Table I

		FORECAST		
		y_1	y_2	Total
Observed	w_1	7	2	9
	w_2	11	80	91
Total		18	82	100

The entry in any column and row gives the number of days out of 100 in which a specific forecast was made, and a specific weather state occurred. From this table we will subsequently derive the distribution $\pi(y, w)$ for this example.

Instead of proceeding directly to find the value to the trucking company of the forecasts just described, let us set up the "protect - don't protect" problem more generally, find a general solution, and then substitute in the specific numbers that relate to our example.

The following notation will prove convenient. Let

π_1 = the relative frequency of forecast y_1, the forecast of unfavorable weather.

π_2 = $1 - \pi_1$ = the relative frequency of forecast y_2, the forecast of favorable weather.

$\tilde{\pi}_{11}$ = the conditional probability of unfavorable weather, given that forecast y_1 is made.

$\tilde{\pi}_{21}$ = $1 - \tilde{\pi}_{11}$ = the conditional probability of favorable weather, given that forecast y_1 is made.

$\tilde{\pi}_{12}$ = the conditional probability of unfavorable weather, given that forecast y_2 is made.

π_{22} = $1 - \pi_{12}$ = the conditional probability of favorable weather, given that forecast y_2 is made.

Thus in the π_{ij}'s above, the first subscript refers to the weather state while the second refers to the forecast. Note that these are <u>conditional</u> probabilities, not joint probabilities.

We now determine the value of an information structure characterized by the above probabilities when the payoff function $\nu(a, w)$ is of the type displayed in the following table:

Table II

WEATHER

		y_1	y_2
Action	a_1	$\nu(a_1, w_1) = -C$	$\nu(a_1, w_2) = -C$
	a_2	$\nu(a_2, w_1) = -L$	$\nu(a_2, w_2) = 0$

Thus C is the cost of protection and L the loss incurred if adverse weather occurs and no protective action is taken. We naturally assume $C < L$, since otherwise the problem is completely trivial.

It is convenient in this problem to work in terms of minimizing cost (disutility) rather than maximizing payoff (utility). So we set $\gamma(a, w) = -\nu(a, w)$. If the decision maker receives forecast y_1, his expected cost if he chooses action a_1 is C, while his expected cost if he chooses a_2 is π_{11} L. Hence

$$\hat{\alpha}(y_1) = \begin{cases} a_1 & \text{if } C < \pi_{11} L \\ a_1 \text{ or } a_2 & \text{if } C = \pi_{11} L \\ a_2 & \text{if } C > \pi_{11} L \end{cases}$$

and $E(\hat{\gamma}|y_1) = \text{Min}(C, \pi_{11} L)$

Similarly, when he receives forecast y_2, he chooses a_1 or a_2 depending on whether C or π_{12} L is smaller:

$$\hat{\alpha}(y_2) = \begin{cases} a_1 & \text{if } C < \pi_{12} L \\ a_1 \text{ or } a_2 & \text{if } C = \pi_{12} L \\ a_2 & \text{if } C > \pi_{12} L \end{cases}$$

and $E(\hat{\gamma}|y_2) = \text{Min}(C, \pi_{12} L)$

The equations above determine the decision maker's best decision rule, and the expected minimized cost for each of the two forecasts. The overall expected cost under the best decision rule is given by

$$E(\hat{\gamma}) = \pi_1 \text{ Min}(C, \pi_{11} L) + \pi_2 \text{ Min}(C, \pi_{12} L)$$

Let us look at the problem from another point of view. There are four decision rules connecting the two messages with the two possible actions. These are displayed in Table III.

Table III

FORECAST

		y_1	y_2
Decision Rule	α_1	a_1	a_1
	α_2	a_1	a_2
	α_3	a_2	a_2
	α_4	a_2	a_1

Of these four rules, the last may be ruled out by an assumption that the forecasts are correctly labeled. Thus rain is not more likely after a no rain forecast than after a rain forecast.

For the remaining possibilities,

$$\hat{\alpha} = \begin{cases} \alpha_1 & \text{if } C < \pi_{11} L \text{ and } C < \pi_{12} L \\ & \quad \text{i.e., if } \frac{C}{L} < \pi_{12} \; (< \pi_{11}) \\ \alpha_2 & \text{if } \pi_{12} < \frac{C}{L} < \pi_{11} \\ \alpha_3 & \text{if } \pi_{12} < \pi_{11} < \frac{C}{L} \end{cases}$$

The climatological probability of unfavorable weather is $\pi^1 = \pi_1 \pi_{11} + \pi_2 \pi_{12}$, which is obviously between π_{12} and π_{11}. Hence α_1 is the best rule on the basis of climatological information alone if $\frac{C}{L} < \pi^1$ and α_3 is the best rule on the basis of climatological

information alone if $\frac{C}{L} > \pi^1$. Thus it is only in the case $\pi_{12} < \frac{C}{L} < \pi_{11}$ that the optimal response to the forecasts differs from the optimal action that would be taken on the basis of climatological information alone, and it is only in this case that the forecasts can have any value to the decision maker. This condition means that unless the weather forecaster is able to better climatology by a certain finite amount (determined by $\frac{C}{L}$) his forecast will have no value at all.

Assuming the forecasts are good enough to have some value

$$E(\hat{\gamma}) = \pi_1 C + \pi_2 \pi_{12} L$$

To determine the value of the information structure which provides forecasts characterized by the indicated probabilities, we must compare this expected cost with the expected cost which would result if climatological information alone were available. The latter is Min $(C, \pi^1 L)$; the best action is to protect all the time if $\frac{C}{L} < \pi^1$, and to run the risk of a loss L occurring if $\frac{C}{L} > \pi^1$. Thus, the value of the information structure -- the saving in expected cost that the forecast makes possible -- is given by

$$V(\eta) = \text{Min}\,(C, \pi^1 L) - \pi_1 C - \pi_2 \pi_{12} L$$

or

$$V(\eta) = \begin{cases} \pi_2 (C - \pi_{12} L) & \text{if } \frac{C}{L} \leq \pi^1 \\ \pi_1 (\pi_{11} L - C) & \text{if } \frac{C}{L} \geq \pi^1 \end{cases}$$

To apply this result to our trucking example, we use the experience recorded in Table I as the basis for estimates of the various probabilities. We find

$$\pi_1 = \frac{18}{100} \qquad \pi_{11} = \frac{7}{18} \approx .389$$

$$\pi_2 = \frac{82}{100} \qquad \pi_{12} = \frac{2}{82} \approx .0244$$

and

$$\pi^1 = \frac{9}{100}$$

We also have $\frac{C}{L} = \frac{20}{500} = .04$. Since the condition $\pi_{12} < \frac{C}{L} < \pi_{11}$ is satisfied, the trucks will be tarped if and only if the forecast is for rain, and the forecasts have value. Since the climatological probability of rain exceeds $\frac{C}{L}$, the best action in the absence of forecasts would be to protect all the time, and the value of the forecasts is given by $\pi_2(C - \pi_{12} L)$, or \$6.40 per truck per day.

We have chosen the value of climatological information as the zero point in our value scale for information structures; but it is also of interest to compare a particular information structure with a hypothetical perfect forecasting system. (The possibility of achieving such a system obviously depends on the probability distribution ϕ as well as on the function η.) In our example, perfect forecasting would mean that favorable weather would be forecast $1 - \pi^1 = .91$ of the time, and the forecasts would always be right, so $\pi_{12} = 0$. Therefore the value of perfect forecasts would be .91 C, or \$18.20 per truck per day.

III. THE MAKING OF FORECASTS: OPTIMAL INFORMATION STRUCTURES

We now turn to the problem of determining how the weather forecaster should summarize the information in his observations if he wishes to make the decision maker's expected payoff as large as possible. That is, taking as given the observations z and their joint distribution ϕ

with the true state of the world, we wish to determine the information structure η that has the largest possible value, $V(\eta)$.

Much of the forecaster's knowledge of present and future weather states relates to weather variables which have no direct relevance to economic activity. For example, few if any activities are directly affected by atmospheric pressure. A decision maker whose best decision depends on the amount of rain has no interest in the atmospheric pressure except insofar as it provides information about the probability distribution of the amount of rain. Not only is much of the forecaster's information irrelevant from the point of view of the decision maker -- much of it is incomprehensible (to raise a problem in semantics we will examine in more detail later). Thus, not only on grounds of cost saving (assuming it is less costly to send less information than more) but on grounds of decoding capacity of the receiver, the weather forecaster must summarize his information, and, in general, translate it.

We shall consider here the problem of determining an optimal information structure for the simple protect - don't protect problem discussed in the previous section. Turning to ϕ, let p(z) be the probability that "unfavorable" weather will occur the day after the forecaster's observation is z. It is clear that (abstracting from the decoding or semantics problem) if the information structure is the identity information structure η^* (all of the forecaster's information is sent on to the decision maker), the optimal decision rule is as follows:[17]

---

[17] In the following equation and elsewhere there is the minor problem that if the decision rule (or information structure is to be a <u>function</u>

$$\hat{\alpha}(\eta^*(z)) = \begin{cases} a_1 & \text{if } C \leq p(z)\, L \\ a_2 & \text{if } C \geq p(z)\, L \end{cases}$$

Further, any information structure η which has the property that each observation z is associated with a unique message y differs only superficially (again abstracting from semantics) from the identity information structure. Any such information structure obviously is sufficient; it permits the decision maker the same expected payoff as the identity information structure. But it also may be very costly.

But, as was stated earlier, a sufficient information structure need not make any distinctions among observations which do not have different implications for the decision maker's action. For the protection problem, the information structure

$$\hat{\eta}(z) = \begin{cases} y_1 \text{ ("unfavorable")} & \text{if } p(z) \geq \frac{C}{L} \\ y_2 \text{ ("favorable")} & \text{if } p(z) \leq \frac{C}{L} \end{cases}$$

is sufficient. It will have positive value provided that the function

---

some arbitrary choices must be made to resolve the borderline cases; in this equation, the case C = p(z) L. We will not bother to treat this minor difficulty with precision, on the grounds that the returns from such a change in the information structure we are providing to the reader would not cover the costs.

p(z) is not everywhere greater than, or everywhere less than, the critical ratio $\frac{C}{L}$ (in which case either y_1 or y_2 could never be sent). Thus for this particular problem an information structure involving only two different messages can be as good as any information structure, in particular as good as the (possibly very costly) identity structure.

We may illustrate these ideas by applying them to the determination of the minimum-message sufficient information structure for the truck protection problem considered earlier. As the meteorological input to the analysis, we take J. C. Thompson's objective scheme for forecasting rainfall in the Los Angeles area.[18] In this scheme, the implications for the probability distribution of rain of several different meteorological observations are summarized in the value of a single variable, designated Y_2 in the Thompson article but identified here as z. The first two columns of Table IV give the absolute frequency distribution for z, and the cumulative relative frequency distribution derived from it. The third column gives the number of days that rain in excess of .15 inches fell after the z value in the indicated interval was observed, and the fourth column is column three divided by column one -- the conditional relative frequency of rain for z observations in the indicated interval. As can be seen, the greater the

[18] J. C. Thompson, "A Numerical Method for Forecasting Rainfall in the Los Angeles Area," Monthly Weather Review, Vol. 78 (July 1950), pp. 113-24. Table IV was obtained by counting points in the scatter diagram, Fig. 5, p. 118 of Thompson's article.

Table IV

z Interval[1]	Number of days	Cumulative relative frequency[2]	Days of rain (rain>.15") for Z in interval	Relative frequency of rain; given Z in interval
Less than 0	54	.151	0	.00
0 to 10	55	.304	0	.00
10 to 20	60	.472	0	.00
20 to 30	45	.598	1	.02
30 to 40	54	.749	3	.06
40 to 50	36	.849	3	.08
50 to 60	21	.907	6	.29
60 to 70	14	.947	4	.36
70 to 80	9	.972	6	.67
Over 80	10	1.000	8	.80
Total	358		32	(.09)

[1]Intervals are inclusive of their upper bounds, exclusive of their lower bounds.

[2]To upper bound of indicated interval.

observed z, the greater the probability of rain.

In Figures I and II, smooth curves have been fitted (freehand) to the relative frequency data. We will interpret these curves as giving true "population" values of the probabilities, since our framework has no place for the problems raised by imperfect knowledge of the basic probability distribution ϕ.

The curve for p(z) in Figure II indicates that a probability of $\frac{20}{500} = .04 = \frac{C}{L}$ is associated with a z value of about 31. Hence a sufficient, minimum-message information structure for the truck protection problem would be to forecast "rain" whenever z exceeds 31 and to forecast "no rain" otherwise. From Figure I, values of z giving rise to the no rain forecast would arise about .63 of the time, so $\tilde{\pi}_2 = .63$. Figure III shows the conditional probabilities of rain given that z is less than any particular "cutoff value," z*. For z* = 31, this probability is about .006, so $\tilde{\pi}_{12} = .006$. Since the climatological probability of rain is .09, the value of this sufficient information structure is given by

$$V(\eta) = \tilde{\pi}_2 \, (C - \tilde{\pi}_{12} L)$$

$$= .63 \, [20 - (.006)500] = \$\ 10.40/\text{truck/day}$$

It is interesting to note that the problem of determining a sufficient information structure and its value can, in this case at least, be cast into a transformation curve -- indifference curve framework. We may consider information structures to be characterized by the accuracy of the no rain forecast and the frequency with which it is made.[19]

[19] This is equivalent to characterizing them in terms of the "Type I and Type II errors," $\tilde{\pi}_{12}$ and $\tilde{\pi}_{21}$.

The expression given above for the value of an information structure can be rewritten to read $V = \pi_2 (C - (1 - \pi_{22}) L)$; it then tells us how these two characteristics can be traded off from the point of view of the consumer of the information: For any given value of V, we can plot the combinations of values of π_2 and π_{22} that yield that value of V.

On the "production" side, there is also a possibility of trade-off between the accuracy and the frequency of the forecast. Greater and greater accuracy in the forecast of no rain can be achieved by insisting on more and more unequivocal evidence of fair weather, but this unequivocal evidence is available less frequently, and consequently the forecast can be made less often. To put the matter in different terms, the forecaster can obviously take any cutoff value z* and choose to forecast no rain if z is less than that value. For any z* so chosen, Figure I shows the frequency with which the no rain forecast will be made, and one minus the ordinate in Figure III gives the accuracy. Plotting the accuracy -- frequency pairs for all possible values of z* gives us the transformation curve between accuracy and frequency.[20]

Figure IV shows the resulting transformation and indifference curves. The optimum previously identified is now associated, of course, with tangency between the indifference curves and the transformation curve. The actual information structure discussed in Section II is seen to be on the transformation curve (the forecaster either is using the Thompson framework or something as good), but the "optimal" η is more than 60 per cent more "valuable" than the actual η. The actual η utilizes a lower accuracy of the no rain forecast than is optimal. Nevertheless, that accuracy is significantly higher than would be achieved if the forecast did not "discriminate" in favor of forecasting rain in

[20]It might be noted here that an improvement in meteorological observation and analysis, an improvement in ϕ, would shift the transformation curve out. Choice among alternative ϕ would have to be made by repeating this analysis for each one.

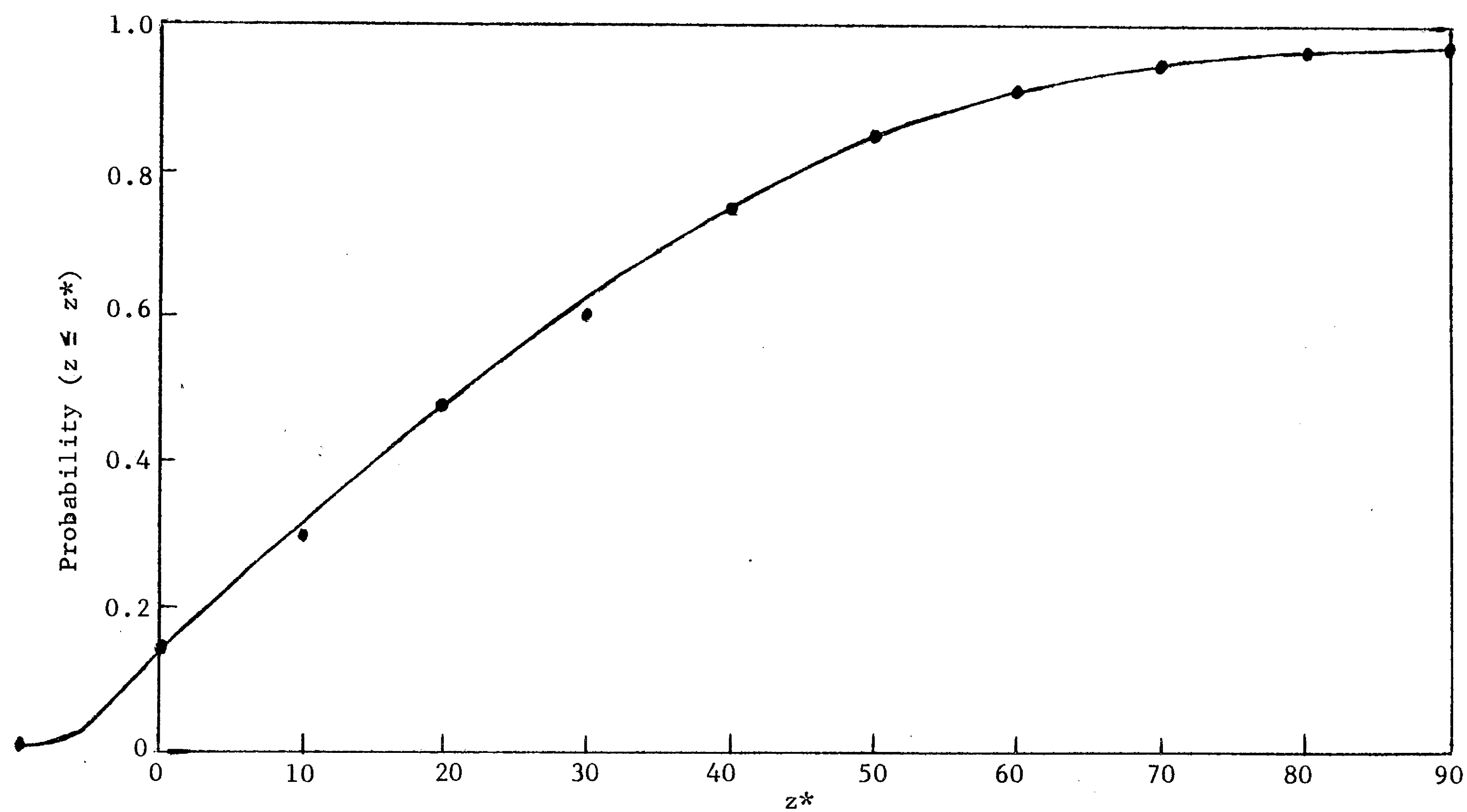
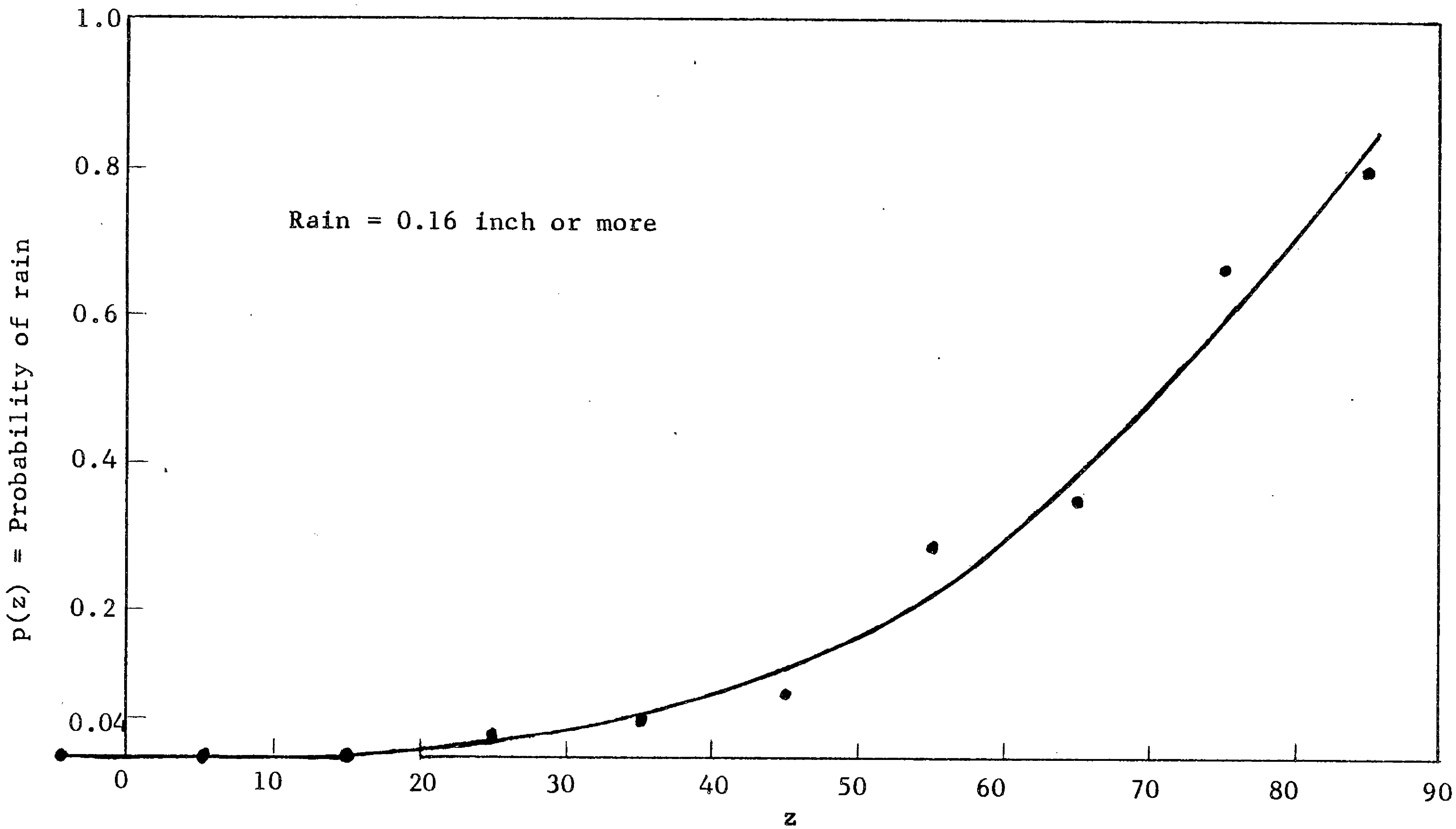


Figure I


1.0
0.8
0.6
0.4
0.2
0.04
p(z) = Probability of rain
Rain = 0.16 inch or more
0
10
20
30
40
50
60
70
80
90
z
Figure II

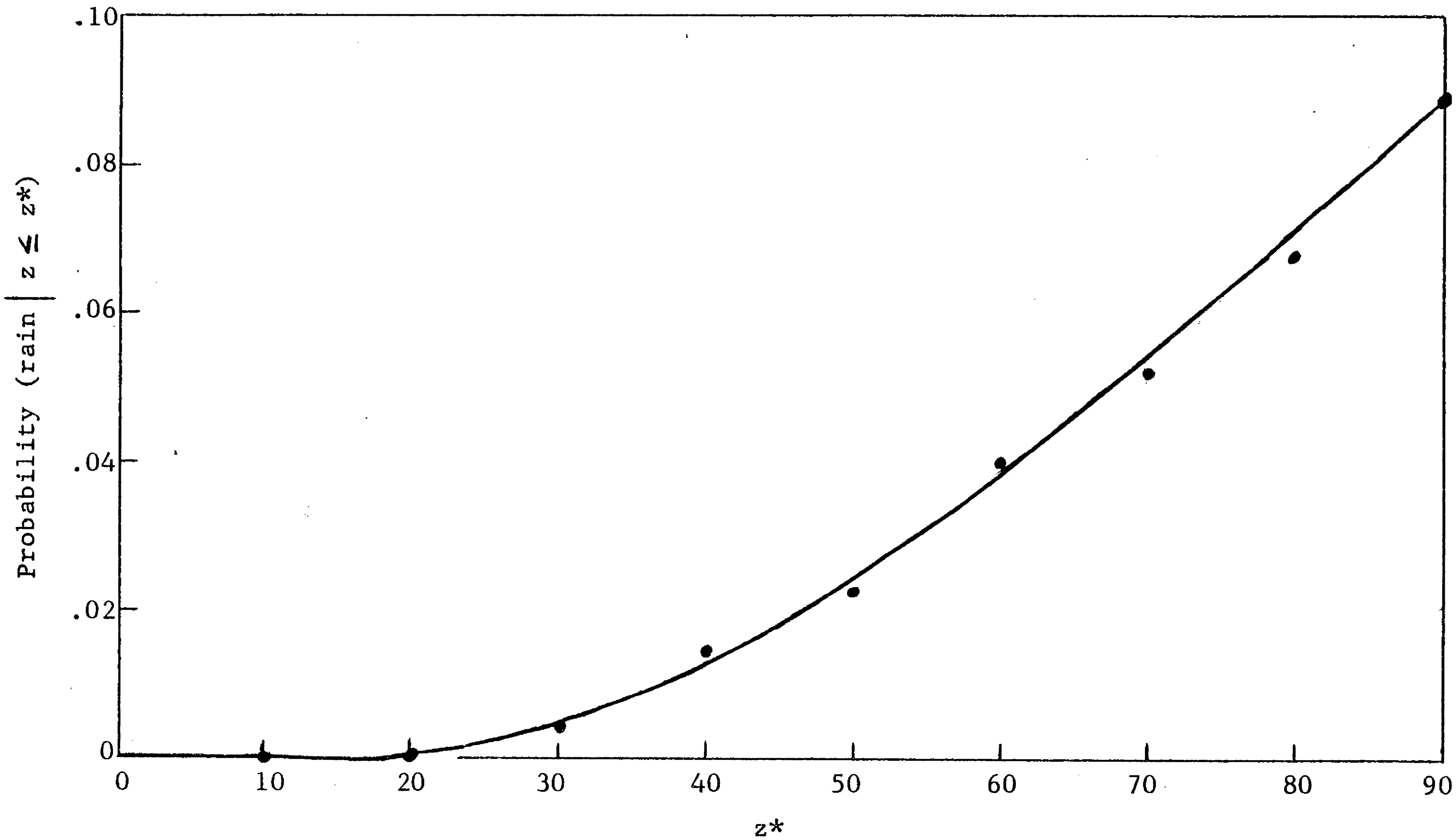
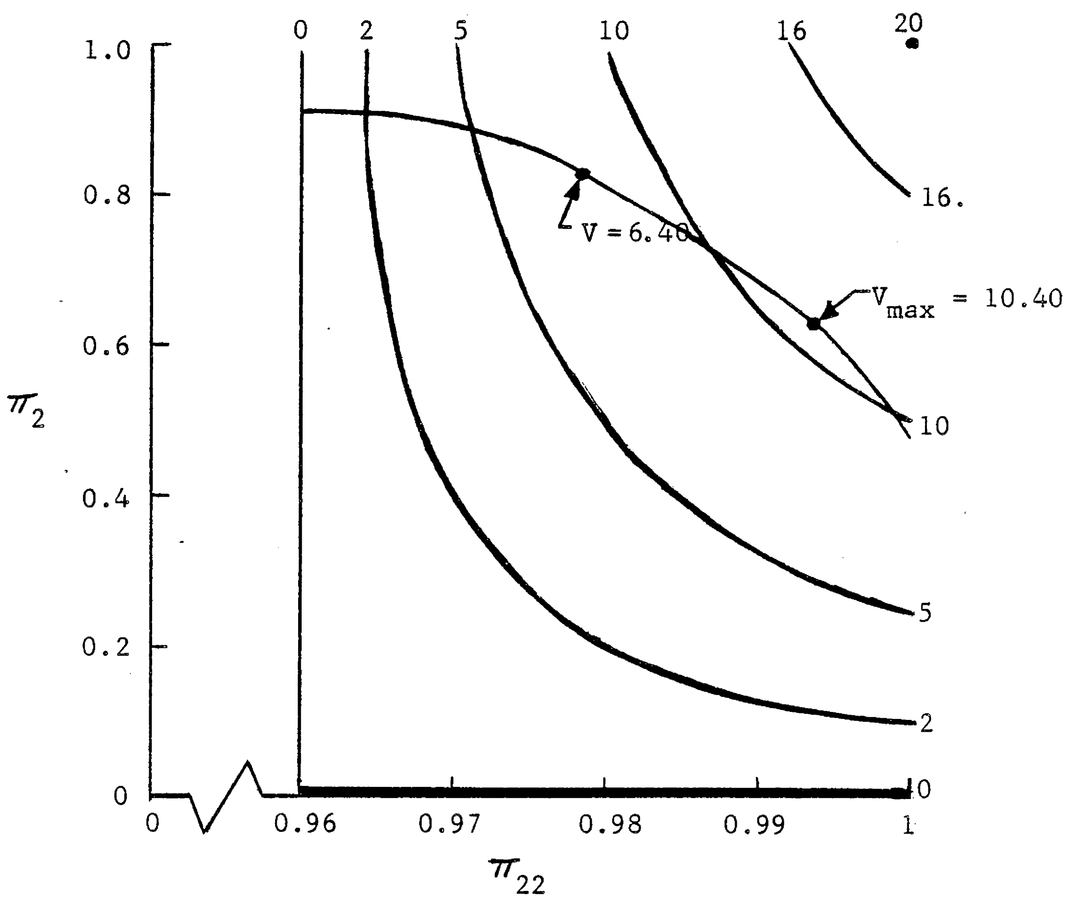


Figure III



1.0
0.8
0.6
0.4
0.2
0
π_2
0
2
5
10
16
20
16.
V = 6.40
V_{max} = 10.40
10
5
2
0
0
0.96
0.97
0.98
0.99
1
π_{22}

Figure IV

doubtful cases. If the forecaster forecast the weather most likely to occur (so that $p(z^*) = .5$), the accuracy of the no rain forecast would be only .95, and the forecasts would have no value at all.

IV. OTHER PROBLEMS OF COST

Our concern with "minimum message sufficient information structures" stems from the conjecture that the greater the number of messages, the greater the cost. However, it also must be recognized that other aspects of the information structure may involve costs, and that cost considerations may affect the optimum information structure in various ways.

For example, let us assume that the cost of an information structure is proportional to the number of signals actually sent (not the number of possible messages). Any signal may be interpreted as a message and no signal also may be interpreted as a message. Let s be the cost of a signal. Then, in the protection problem the expected cost of the information structure, $K(\eta)$, is given by

$$K(\eta) = s \text{ Min } (\pi_1, \pi_2)$$

If the cost function has this form, the minimum message sufficient information structure discussed above is not generally optimal in the sense of maximizing $R(\eta) = V(\eta) - K(\eta)$. In fact, it is optimal only in the case $s = 0$, when information structures involving more than the minimum number of messages are also optimal.

To see the reason for this conclusion, consider the sufficient information structure discussed above, where $\pi_1 = .37$ and $\pi_2 = .63$. Obviously the message y_2 (no rain) should be conveyed by "no signal."

But now consider a value of z such that p(z) exceeds very slightly the critical ratio $\frac{C}{L}$ = .04. For such a value of z, it makes almost no difference which action the decision maker takes. And if p(z) is close enough to .04, the improvement the signal permits in the decision maker's expected payoff will not be sufficient to cover the cost s of sending the signal. Thus the forecaster should send the rain forecast somewhat less often when there is a positive cost associated with sending it than is required by the sufficient information structure.

Under these circumstances, the forecaster should actually choose the value of z such that $p(z) = \frac{C + s}{L}$ as his cutoff value, forecasting rain only if z exceeds this value. For whenever the rain forecast is made, both the protection cost C and the signal cost s will be incurred with certainty, while if the no rain forecast is made the expected cost is still p(z) L.[21]

More complicated problems of cost arise when the decision maker has more than two possible actions. These problems, and others as well, face a central weather bureau which serves several different decision makers with the same set of signals and we shall examine them here in this broader context. These broader problems involve both difficulties with respect to institutional arrangements, and with respect to semantics.

To an increasing extent, large companies whose operations are affected by the weather are hiring private meteorological firms to

---

[21] It is the other way around, of course, if the rain forecast is made more than half the time; the cutoff value should then be the z such that $p(z) = \frac{C - s}{L}$. In the general case, it is necessary to make an explicit comparison of the two alternatives $p(z^*) = \frac{C + s}{L}$ (no signal = no rain), and $p(z^*) = \frac{C - s}{L}$ (no signal = rain) to determine which is better.

forecast the weather for them, rather than relying on the services of the Weather Bureau.[22] These firms rely in general upon the data obtained by the Weather Bureau's network of reporting stations, and they typically do not rest the case for the value of their services upon an alleged superiority in scientific skill relative to the Weather Bureau. What they do offer is, in our terms, a superior information structure: they provide more information about the weather dimensions which affect the client's operations; they concentrate on forecasting events in the space and time frame of interest to the client; and, it appears that they sometimes take into account the risks associated with the alternative actions open to the decision maker and "slant" their forecasting accordingly. It is interesting, therefore, to examine the question of what the appropriate division of labor is between a public information source such as the Weather Bureau and the private firms which supplement the services it provides to particular decision makers.

Let us continue, for the moment, to assume that the only relevant weather states from the point of view of any decision maker are rain and no rain. But any decision maker may have m possible choices of degrees of protective action. If the probability of rain is high it may be optimal to take quite expensive protection; if there is little

[22] See the article, "How's the Weather," The Wall Street Journal, March 19, 1963, p. 1, where it is reported that private weather forecasters are expected to do $15 million worth of business in 1963 as against $10 million in 1961, and that the number of private meteorological firms in the United States is 35, up from 27 two years ago.

chance of rain, less protection. Given m possible actions, we know that each may be optimal for a given range of rain probability, and thus to provide sufficient information to the decision maker may require as many as m possible messages. If there are N decision makers, all of whom are to receive the forecast, as many as $\sum_{k=1}^{N} m_k$ signals may be required for a sufficient information structure. Clearly, this might be very costly and, if the cost of the structure were dependent upon the number of possible forecasts, undoubtedly a system involving less than sufficient information would be optimal.

Clearly; this is the type of situation where private weather consultants, drawing on weather bureau basic information, but tailoring their forecasts to the needs of a particular client, can play a useful role. This institutional arrangement makes economic sense if the major problem in having a central weather bureau provide information of sufficient detail to be useful to all lay in the costs of having an information structure with such a large number of possible messages.

However, in many instances it would not appear that the number of possible messages has much to do with the transmission costs of the information structure. The problem of a many message structure, at least in the weather forecasting case, lies in the added difficulty of decoding or understanding the messages, rather than in the costs of sending them. In the example we have studied the problem is one of semantics, not of channel capacity. We have not until now attached significance to the exact words or numbers used to convey the signal to the decision maker. We have assumed that the decision maker knows the true relation between the signals he receives and the probabilities of the weather states,

presumably through long experience with the forecasts. But the question of how the labeling of the forecasts may affect their value cannot be pushed aside in a discussion of how the forecaster's knowledge of the weather should be conveyed to decision makers. Intelligibility is obviously an important consideration. Or, to put it another way, we must not ignore the costs of information processing which the labeling of the forecast imposes upon the decision maker, or the possibility that the decision maker may be unaware of the relevance of the signals to his problem if the labeling is inconvenient.

If the number of decision makers served by the forecasts and the number of actions open to each are small, or if the decision problems are virtually identical, it may be possible to construct a suggestive set of verbal labels for the small number of forecasts required for a sufficient information structure -- slight chance of rain, moderate chance of rain, high chance of rain, etc. But the most obvious practical solution to the labeling problem in this case is for the forecaster simply to provide the probability p(z) itself, to whatever degree of accuracy may be warranted by the state of knowledge of the joint distribution of z and w.[23] A major advantage of this solution is that the forecaster does not need to have detailed knowledge of all of the decision problems in order to provide a sufficient information structure (or a close approximation thereto).

As compared with the more typical alternatives of providing some

---

[23] Here we must recognize as a practical matter that knowledge of the distribution ϕ is less than perfect. Clearly, if p(z) can be any real number in the 0 - 1 interval, the task of communicating its precise value is going to be formidable.

"categorical" forecast, or a vague verbal description of the forecaster's uncertainty, probability forecasts offer a much closer approximation to sufficiency. The benefits from this can be very large when, as in the case of the Thompson objective scheme, different observations provide different degrees of certainty that certain weather states will or will not occur. Much of the forecaster's information about future weather is kept from the decision maker when high certainty situations are lumped with much more doubtful situations in an information structure which involves a small number of distinct signals. For example, a system of forecasts which merely described the weather state most likely to occur involves such a drastic summarization of the forecaster's information as to wipe out, for many decision makers, the information value potentially derivable from the forecaster's observations. If some of these decision makers then turn to private forecasters who charge a positive price to supply information which they do not produce and which has a very small transmission cost, only a part of the social loss resulting from over-summarization can be recouped.

Forecasts of rain stated in probability terms are now standard in several American cities, and our discussion above suggests that the extension of this practice is desirable. However, this does not begin to answer the question of how the public forecaster should operate.

While better semantics on the part of the Weather Bureau would help to reduce the need for tailored forecasts, it is clear that in many instances decision makers do require specialized forecasts. For the Weather Bureau to provide close to sufficient information to all would be very costly, even if the semantics were good. While many decision problems can be adequately treated by distinguishing only "rain" and "no rain" as weather states, for other problems the decision maker will want to

know a great deal more about the probability distribution of amounts of rain than the probability concentrated at the origin. Still other problems depend on events in other weather dimensions -- wind and temperature, type of precipitation, and so forth -- and different decision makers will in general be interested in events in different localities. When it is recognized that taking action is generally time consuming and that predictions of the weather at a future time t will improve as t gets closer, it becomes clear that many decision makers face problems whose solution depends in principle upon the joint distribution of weather states and forecasts over time. Without violence to the complexity of real world decision problems, the amount of detail that a sufficient-for-everybody information structure would provide can be regarded as essentially infinite. Therefore, constraints on communication and information processing cannot be disregarded indefinitely. The public forecaster <u>must</u> provide an information structure that is less than sufficient for some decision makers, and the question is what amounts and types of detail it is the function of the public forecaster to provide, and what should be left to private forecasters.

Clearly, the answer is that the public forecaster's choice of an information structure should reflect the importance of the different weather dimensions as determined by the number and types of decision problems faced and the sensitivity of the payoff to the action taken, and in those dimensions the information provided about the probability distribution should reflect any clustering of "critical points" that may exist -- whether these are critical points for probabilities, for the expectations of certain quantities, for the time that will elapse before certain events occur, or something else. But it is equally clear that

considerable progress will have to be made in the economic theory of information before the vague proposition just stated can be made precise.