

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut

COWLES FOUNDATION DISCUSSION PAPER NO. 280

Note: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than mere acknowledgement by a writer that he has access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

ECONOMIC ASPECTS OF INFORMATION

Luigi M. Tomasini

August 20, 1969

ECONOMIC ASPECTS OF INFORMATION

by

Luigi M. Tomasini^{*}

1.1 Let us start by giving a definition of expectations. In Georgescu-Rogen's [2] words "expectations is the state of the mind of a given individual with respect to an assertion, a coming event, or other matter on which absolute knowledge does not necessarily exist".¹ In other words, we are in a situation where the individual (or group) is in a state such that the consequences of a given action are partially known. Drinking my coke when I am thirsty can give me a pleasant feeling of satisfaction or can cause me problems (stomach-ache, for example). In this way, whenever I am offered a coke, I will have to face the problem of a pleasant feeling or a stomach-ache. This last proposition suggests the possibility of comparing expectations and therefore of measuring them. The dispute between the Classical group which later developed in to the objectivist school on the one hand, and the subjectivist on the other, is concerned with this possibility.

* I wish to thank William Baumol and Herbert Scarf who at different stages contributed to improve a first draft of the paper. Responsibility is, of course, entirely mine. The research described in this paper was carried out under grants from the National Science Foundation and from the Ford Foundation.

¹ Georgescu-Rogen adapts a definition of probability given by De Morgan. Cfr. [2].

In our analysis we will accept entirely the argument supported by the subjectivist school, i.e. that probability expectation is a numerical coefficient which measures the subjective degree of belief that a given event will happen. The numerical coefficient is built with the aid of the betting quotient any subject is willing to accept on a given uncertain event, the betting quotient being expressed in terms of utility. In this formulation, different individuals will have different betting quotients, i.e., they will attribute different coefficients to the probability P of a given event.

1.2 Many economists continue to use the distinction between risk and uncertainty developed since Knight's famous work. The difference, according to these authors lies in the fact that it is possible to assign mathematical probabilities to risk but not to uncertainty. Where this distinction is relevant and the decision-maker faces risk, i.e., events which are expected to occur and for which we have a history that provides numerical coefficients, the mathematical (objective) probabilities should clearly be used. In the case of uncertainty he should choose a Principium which depends on his personal a priori probabilities. These principia, such as the safety margin (Fellner, 1948), the index of pessimism and optimism (Hurwicz, 1951), the minimum subjective loss (Savage, 1951), the minimum chance of disaster (Roy, 1952 and 1956) are however a mixture of subjective and rational (objective) elements. It seems worth emphasizing that once a principium is chosen any pay-off matrix will be the same in form as that under risk.

The difference between risk expectations and uncertainty expectations is, in any event, unnecessary as it is always possible to obtain a

betting quotient from any individual for any outcome.² This implies that all expectations are measurable and that no meaningful distinction can in real life be made between risk and uncertainty. From now on when we talk about uncertainty we will consider this term to include the actuarial risk of any outcome.

In real life the betting coefficient will be the result of the quantity of data (information) drawn from past experience, and the ability to assign to the latter a weight based on an intuitive judgment. Experience in this case will supply information that can help the individual to face and forecast an event. When this experience is unavailable or strongly limited (qualitatively or quantitatively) the decision-maker will face the future formulating a decision based on his own judgment.

1.3 The decision-maker is assumed to be adapting continuously to a situation of uncertainty. In this adaptive process he formulates decisions and takes actions. Given the decision-maker's tastes and resources his decision (or strategy) will depend on the probabilities he attaches to events which will determine the result of each action. These probabilities depend on the level of information available at the time he formulates his decisions. In fact, a quantitatively or qualitatively (better) different level of information at time t_1 can change the weight given by the decision-maker to states of

²One could ask an individual what probability (betting quotient) the Americans have in his judgment of getting to Mars by the year 1980.

nature and can also reveal the feasibility and result of actions he did not consider at time t_0 .

An information system is defined, à la Marschak, as a set of potential messages which the decision-maker will use by responding to each message with some action. In this definition two elements need to be emphasized for our analysis. First, we consider information as the output, while raw data or facts constitute the input. A second point is the fact that information is considered as a process which continues with the passage of time. The adaptive process of the decision-maker to uncertainty assumes this form: Information \rightarrow Decision \rightarrow Action \rightarrow Information. An action is considered optimal with respect to information, if the result of the action could not be, on the average, improved by choosing any other available action. The value of a given information system is measured by the utility achieved in facing expectations. Given the "tastes and resources" of the decision-maker, it is possible to measure the value of any information system.

It is obvious that information is costly and that its cost has to be considered before the decision maker formulates a strategy. If I do not have a TV to look at the weather forecast and I have to buy one in order to decide if tomorrow I have to take my umbrella, even if the value (gross value) of TV weather forecast information is greater than my looking at the sky, the net value will be less. It is this latter that the decision maker will consider before taking decisions.

Another cost is that of thinking, (Cfr. [6]) i.e., the cost in terms of efforts that the decision maker supports selecting an action which is optimal for the information received. The net utility of an action will

be obtained by subtracting from the value of information the cost of obtaining it (in a general sense) plus the cost of decision making.

1.4 An example from the theory of the firm can illustrate the nature of the problem. According to the technique of marginal analysis any input is hired in a productive process up to the point where the value of its marginal product equals its price. In this framework the value of information can be given by the gains or losses in money terms, *ceteris paribus*, the firm could obtain depending on the availability or lack of information.

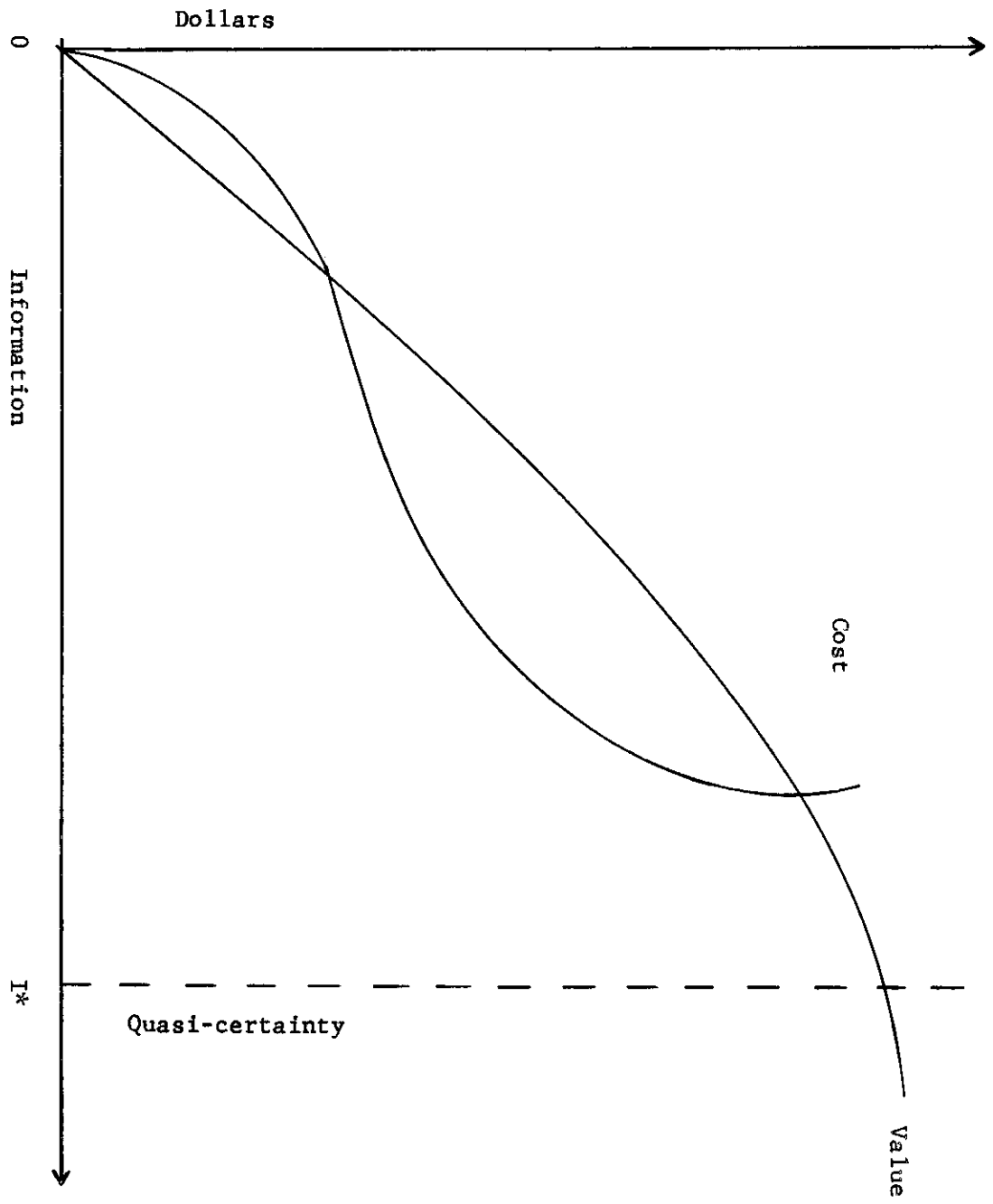
At this point it is useful to introduce two more assumptions:

(i) more information will reduce the degree of uncertainty and; (ii) there will be a limit to the level (quantity and/or quality) of information the firm will purchase, the limit being determined where any further unit of information will not increase the knowledge of the future. We will call this a state of "quasi-certainty."

It is possible to provide a graphical representation of the problem. In Fig. 1 we represent "quantity" of information on the abscissa, and the cost of obtaining information and its value--both in monetary terms--on the ordinate.

The behavior of the cost curve in the first part is explained by considering the fixed costs (employees, computers, surveys, transmission of informations, etc.) the firm has to face. The optimal level of information will be I_0 where the marginal cost equals the marginal value of information.

The value of this type of analysis is however extremely limited. In fact, it has been long recognized that marginal analysis is of limited value where we introduce time into the model. In our particular case there are two interrelated problems which make it inappropriate: (i) information as defined here implies an essentially dynamic world and moreover a world which is



continuously adapting; (ii) marginal analysis cannot handle imperfect market situations.

It seems obvious that in order to have even a slight understanding of the problem we have to consider only dynamic processes and among these only those which permit us to take account of reactions of decision makers to changes in the amount of information available to them.

2.1 To have a better understanding of the problem sketched in the previous sections, it is useful to formalize the concepts we have discussed.

Consider:

$A = \{a_1, \dots, a_n\}$: a set of feasible actions;

$Z = \{z_1, \dots, z_n\}$: a set of states of the world;

$X = \{x_1, \dots, x_n\}$: a partition set of Z of payoff-relevant events³ (cfr. [4];

$\rho(x, a) \rightarrow R^l$: a payoff function (an ordinal utility function which respects the axioms of the theory of choices);

$p(x)$: a (personal probability function such that $p(x) \geq 0$ all x ; $p(x \cup x') = p(x) + p(x')$ if $x \cap x'$ are disjoint; $p(Z) = 1$;

$Y = \{y_1, \dots, y_n\}$: a partition set of Z of available information.

In the case of certainty the decision maker knows x . By the rationality assumption he will then choose an action $a^* \in A$ so as to

$$(2.1) \quad \max_{a \in A} \rho(x, a) = \rho(x, a^*) \geq \rho(x, a) \text{ all } a \in A .$$

³ The set $V = \{v_1, \dots, v_u\}$ is said to be a partition of a set Z if: (i) $v_i \subset Z$ ($i=1, \dots, u$); (ii) $v_i \cap v_j = \emptyset$ for $i \neq j$; (iii) $v_1 \cup v_2 \cup \dots \cup v_u = Z$.

Definition 1. A decision-function (or strategy) is a mapping d that associates each state of the world $x \in X$ with some action $a \in A$, i.e.

$d : X \rightarrow A$.

Definition 2. An optimal decision function d^* associates each $x \in X$ with an optimal $a^* \in A$.

In the case of uncertainty, the rational decision maker will choose an action $a^* \in A$ only if

$$\sum_{x \in X} p(x) \rho(x, a^*) \geq \sum_{x \in X} p(x) \rho(x, a), \text{ all } a \in A,$$

and will behave so as to

$$(2.2) \quad \max_{a \in A} \sum_{x \in X} p(x) \rho(x, a).$$

2.2 If we rule out the case of "complete information" which is, of course, indeed an extreme situation, we are left with the problem of choosing an action and of basing such a choice on the knowledge received from some "informative source".

Consider the partition Y of Z which defines the information set. The decision maker because he is ignorant of which $x \in X$ will occur, bases his decision on the knowledge of which $y \in Y$ will occur. Of course, in the case of complete information Y and Z coincide.

The information set Y is noiseless (no errors) with respect to the set of events X if X is finer than Y ($X \text{ f } Y$).⁴

⁴ Given two partitions V and V' of a set Z , if V is a sub-partition of V' we say that V is finer than V' or $V \text{ f } V'$ or, conversely, V' coarser than V .

In such a case there exists a single-valued function $\Psi: X \rightarrow Y$; i.e. each event $x \in X$ is associated with exactly one message $y \in Y$.

Any information set Y lies somewhere between the set of complete information (Y_{\max}) and that of zero information (Y_{\min}). More formally,

$$X = Y_{\max} \text{ f } Y \text{ f } Y_{\min} .$$

Consider now a noisy (with errors) information set. In such a case we do not have a single-valued function Ψ . It is however possible to associate each value $x \in X$ with a probability distribution on Y . In such a case the random function, Ψ takes $x \in X$ into the matrix of conditional probabilities $p(y|x)$ of $y \in Y$, given x .

It is obvious that the noiseless case is a special case of the noisy one. In the case of complete information, for example, Ψ is the identity matrix.

2.3 The probability function $p(x)$ depends on the information set available to the decision maker. His problem (eq. 2.2) can therefore be restated as

$$(2.3) \quad \max_{a \in A} \sum_{x \in X} p(x|y_0) \rho(x, a)$$

where $y_0 \in Y$ indicates a fixed value of the message $y \in Y$.

Given the set Y a decision maker who responds to each observation $y \in Y$ with an optimal action will obtain, averaging over all such observations, a higher (or at least not lower) expected utility than by using any other way of responding to those observations by actions. Marschak calls this (maximum) expected utility the gross value of the information set Y , i.e. $U(Y)$.

We have therefore,

$$(2.4) \quad U(Y) = \sum_{y \in Y} p(y) \max_{a \in A} \sum_{x \in X} p(x|y) \rho(x, a)$$

where $U(Y)$ could be written as $U(\Psi)$ if we neglect observations costs and consider all the messages with the same Ψ .

Definition 3. A decision function d is a mapping which associates each value $y \in Y$ with some action $a \in A$ i.e. $d: Y \rightarrow A$.

Definition 4. An optimal decision function d^* is a mapping which associates each observation $y \in Y$ with an optimal action $a^* \in A$.

The problem for the decision maker is to find an optimal decision rule d^* given the information set Y . If different information sets (Y, Y', Y'', \dots) are available, we have to compare (neglecting for the moment the costs of these different information sets) the values $U(Y), U(Y'), \dots$. More formally we can state that Y is more informative than Y' with respect to X , i.e. $(Y > Y' | X)$ if

$$(2.5) \quad U(Y; \rho_{X \times A}, p_X) \geq U(Y'; \rho_{X \times A}, p_X)$$

for all the ρ defined in the cartesian product $X \times A$, all probability functions p on X .

Marschak proves that, in the case of noiseless information set the eq. (2.5) coincides with the concept of "finer than". More formally, let $X \in Y, S \in Y'$ then $Y > Y' | X$ if and only if $Y \in Y'$. (For the proof cfr. [4]).

2.4 Up to this point the decision-maker problem has been the one of choosing an information set Y with the greatest gross value, i.e. he will act so as to maximize the expected payoff simultaneously with respect to X and to the function $d_Y: Y \rightarrow A$.

Not all information sets are feasible. In fact there will be some which will not be available to the decision maker. The same applies to the decision functions. The unavailability depends on the infinite costs which the decision maker faces. Regarding costs of information (collection, elaboration, reading of data, etc.) we can distinguish following Marschak between fixed and variable costs. The meaningfulness of such a distinction is however questionable as is the one normally used in economics.

An important question is the one which concerns the measurement of the observation costs and the payoff (gross value of information). It is evident that it is rather naive to measure both in terms of money. The problem is somewhat complicated by the fact that in reality we are interested in measuring the net value of information, i.e. the value of gross information minus the costs, including the cost of thinking.

An alternative is to measure gross payoff, observations costs and decision costs in the same unit (utilities). The net value of information (V_d) would therefore be

$$(2.6) \quad V_d = (Y; \rho, P_X, c, \gamma) = U_d(Y, \rho, P_X) - k(Y) - \gamma(d_Y),$$

where k represents the total cost function, γ is a function from the set d_Y of feasible decision functions on Y to the set R^l of utilities.

Sometimes such a decomposition of the set value is not possible, in such a case V_d must be evaluated directly.

2.5 Consider a decision maker who is informed about two different future events: sunshine tomorrow and a crisis in the stock market such that he will lose all his capital. In such a case, although the "amount of information" received by

him (in the form yes-no for the two events) is the same, the "value" of it will be--assuming his rationality--quite different. At the same time it is conceivable that the price he is willing to pay to receive the information regarding the two events is "quite" different.

The simple considerations should make clear that what is meant by the "value" of information in communication theory (Shannon, 1948) is in fact the cost of transmitting a message. As is well known, such a theory studies, considering $Y = X$ (message sent = event), the optimal choice of a communication function $\Gamma: X \rightarrow Y$, usually a stochastic function,

$$(2.7) \quad \Gamma = (\gamma_{xy}) ; \quad \gamma_{xy} = p(y|x) ,$$

where $x \in X$ is the message sent and $y \in Y$ the one received.

The payoff function considered is

$$(2.8) \quad \rho(x,y) = \begin{cases} 0 & \text{if } y = x \\ -1 & \text{if } y \neq x , \end{cases}$$

that is, every error in identifying the message sent has the same negative payoff (penalty).

The problem is to choose among the set $\{\Gamma\}$ of all possible communication systems a Γ that minimizes the "probability" of error for a fixed cost of Γ and obtain for each cost, the most economical communication system.

2.6 The communication function Γ depends on the channel and on the code.

The event $x \in X$ is encoded into a message-input; the channel is characterized by a transmission function τ so that a message-output is received and decoded at the other end of the channel. The problem is then to select, given known costs, the efficient pair $\{\tau, c\}$.

The amount of messages that can be carried by the channel in a given unit of time is called the capacity of the channel. Assuming there is no transmission noise (i.e. that the channel is errorless) we can increase the capacity of the channel by increasing the "length" of a "block", i.e., instead of transmitting a message every unit of time we transmit one (in the same unit of time) about a sequence (block) of two events. As the length of the block increases the needed channel capacity decreases and converges towards the function p_X

$$(2.9) \quad p_X = -\sum_{x \in X} p(x) \log_r p(x),$$

where p_X is usually denoted by $H(X)$. For convenience we shall use logarithms to the base 2 ($r = 2$). If the event $x_i \in X$ is the only possible, $p_i = 1$ and all the other $p_j = 0$ ($i \neq j$). In such a case $p_X = 0$, i.e. it gets its minimum value. For $p_i = 1/n$ for all $i = 1, \dots, n$, $p_X \max = \log_2 n$, i.e. it gets its maximum value.

To sum up, then, one can say the decision maker in the case of uncertainty chooses an action and bases such a choice on the information set available to him. He has to select simultaneously an information set, a communication system and a decision rule. This triple is optimal if there exists no other triple which gives on the average a better result than that obtainable by the one considered. It is important to note that if the choice of each component of the triple is made independently, the decision maker may suffer losses.

3.1 In order to have a better understanding of the decision maker's problem in the face of uncertainty we have to specify the system in which he operates.

Two variables are of fundamental importance for the decision maker: resources and information. Until recently economic theory has dealt with the problem of allocation of resources among competing uses--assuming that information is equally distributed among all the individual components of an economic system. In fact, one usually makes an even stronger assumption i.e. that all individuals have perfect information.

For more realistic models of the economic system we recognize that such an assumption is rather "restrictive". Moreover it is our contention that information plays a more important role than the one economists are willing to attribute to it. In particular, the distribution of the stock of information (knowledge) at a given time determines the flow of information from the environment to the system (whether it is an economy, an individual or a firm) thereby, increasing the stock of information. The resulting process would be cumulative and would have a tendency to maintain and eventually increase differences in the stock of information among the components of the system as well as differences in the flows of information.

A decision maker can, in general, be described by an ordered n.tuple of n (in our case $n = 2 =$ resources and information) economic variables or characteristics. If we define a function for each n.tuple we can identify the number of decision makers who possess, at a given time, the same amount of the two characteristics. The economic state of the system, can be, therefore, represented by n such functions.

The state is not stable. In fact, if we define an ordered pair of n.tuples the first representing the initial state, and the second the subsequent state, we can define a transition probability function which tells us

what the probability is that a given person in a given state will move to a different state in the next period.⁵

3.2 As we have seen all decision makers can be classified according to the amount of information (stock of knowledge) and resources they possess at a given time. For a given state of information we can find a number r of individuals who have the same amount of information.

Consider now the entropy of a system p_X which measures the rate at which information is transmitted by the environment to the system. The stock of information possessed by the i th decision maker at the stage s is measured by

$$p_{X \max} - p_{Xs}^i$$

Of course, if he knows $p_{X \max}$ then the stock of information is maximal.

If $p_{X \max} = p_{Xs}^i$ the stock of information is zero. This stock of information defines the state of knowledge for the i th decision maker. The problem consists mainly in the measurement of p_X , the entropy of the system.

In a given economy neither the resources nor the information (stock) is equally distributed among the individuals. Each decision maker has the possibility to move from one state of information to another, the same is true for resources. The change can be interpreted as a loss of information (due to "forgetting") and as a loss of resources (choice of wrong decision). It is also conceivable that the choice among all the feasible actions is such (optimal) that the decision maker increases at the same time the resources and the stock of information.

⁵ In a deterministic process (transition probability = 1), given the initial state the process is completely determined.

The differences in the stock of information and of resources explain the differences in behavior which can be found among the decision makers. If all of them have the same stock of information and different resources the optimality of an action will be different for different decision makers who face the same event. In the economic literature such behavior is associated with the theory of expected utility and the distinction between risk-takers and risk-lovers. The same sort of reasoning applies in case of inequality in the stock of information and equality of resources among the individuals.

3.3 A decision maker operates by formulating decisions. Let us assume that he has to formulate some decisions at time zero with a zero stock of information. Although such an assumption is open to discussion--as is clear if one considers the Laplace principle--the decision maker will choose an action which, assuming there are n feasible actions has probability $1/n$ of being optimal. At time one a new decision has to be taken. This time however the previous information (experience, knowledge) will increase the value of the probability coefficient of choosing an optimal action. After t trials (or after a "reasonable" amount of time) the decision maker will reach some kind of statistical equilibrium with his environment. The passage of time being associated with the increase in information.

The amount of information which flows from the environment to the decision maker can be increased or decreased according to the quantity of resources which any decision maker allocates to its acquisition. This proposition has its "dual". In fact, it is still true that the increase in the amount of resources depends on the choice of a decision, whose probability of being optimal increases with the increase in the flow of information.

Consider two decision makers who start with the same amount of resources but with different stocks of information. Let us further assume that the rate of growth in resources is a given constant α common to both and that they want to reach a given target (a given stock of resources and information) which we will consider as a steady state.

The length of the path to this target will be different. In fact it will be a function of the initial stock of information and of the flow of information which the decision makers receives with the passage of time. This implies that one who starts with a larger stock of information will reach, *coeteris paribus*, the steady state in a shorter period of time.

The cost--in terms of loss of resources--to reach the steady state will be greater the less information the decision maker possesses. Moreover as we implicitly assumed that both face the same sequence of events in reaching the target, the two paths to the steady state cannot, in general, cross each other. This rules out the possibility of overtaking the hypothesis which is the basis of the modern theory of optimal growth.

Another aspect to be considered is the interaction among all the decision makers in a system; their decisions, in fact, influence each other. In such a process the system has to have an adjustment mechanism which in an economy is provided by a price system. This last can be envisaged as a control device which allows the economy as a whole to reach a "type" of statistical equilibrium.

References

1. Abramson, N. (1963), Information Theory and Coding, New York: McGraw-Hill.
2. Georgescu-Rogen, N. (1958) The Nature of Expectation and Uncertainty, in Bowman, M. J. (ed.) Expectations, Uncertainty and Business Behavior, New York; reprinted in Georgescu-Rogen, N. (1966), Analytical Economics, Harvard University Press.
3. Marschak, J. (1959), Remarks on the Economics of Information, in Contributions to Scientific Research in Management, Western Data Processing Center University of California, Los Angeles.
4. _____ (1963), The Payoff -- Relevant Description of States and Acts, Econometrica, Vol. 31: 719-26.
5. _____ (1964), Problems in Information Economics, in Bonini, C.P. et al. (eds.), Management Controls: New Directions in Basic Research, New York: McGraw-Hill.
6. _____ (1966) Economics of Planning and the Cost of Thinking, Social Research, Vol. 33: 151-9.
7. _____ (1968), Decision Making: Economic Aspects, in Sills, D. L. (ed.) International Encyclopedia of the Social Sciences, Vol. 4, New York: Crowell Collier and McMillan.
8. _____ (1968), Economics of Inquiring, Communicating, Deciding, American Economic Review Papers and Proceedings, Vol. LVIII: 1-18.
9. _____ and Miyasawa, K. (1968), Economic Comparability of Information Systems, International Economic Review, Vol. 9: 137-74.
10. Murphy, E. R. (1965), Adaptive Processes in Economic Systems, New York: Academic Press.
11. Ozaga, S. A. (1965), Expectations in Economic Theory, London: Weidenfeld and Nicolson.
12. Savage, L. J. (1954), The Foundations of Statistics, New York: John Wiley and Sons.
13. Simon, H. A. (1957), Models of Man, New York: John Wiley and Sons.