*CHAPTER 3*

# A NONCONCAVITY IN THE VALUE OF INFORMATION

ROY RADNER
*Bell Laboratories*

and

JOSEPH E. STIGLITZ*
*Princeton University*

## 1. Introduction

It is a truism of statistical decision theory that costless information is harmless, and if it is relevant to the decision problem at hand then it will have positive value.[1] In other words, as an input into economic decision-making, the productivity of information is non-negative, and is typically positive. Does information also obey the Law of Diminishing Returns? In particular, is the marginal productivity of information strictly positive, at least for small amounts?

The marginal productivity of information depends, of course, on the way the quantity of information is measured. We know that, in general, there is no way to measure the quantity of information (as a real number) so that, of two information structures, one is more valuable if and only if it has "more" information.[2] Nevertheless, for a particular decision problem there may be a family of available information structures indexed by a real parameter, in which larger values of the parameter correspond to more costly information, and in which a zero value of the parameter corresponds to a structure that is both costless and completely "noninformative" (in a sense to be defined below). In this context, we shall be able to talk meaningfully about a "small amount of information", i.e. the case in which the parameter is close to zero.

* Research supported in part by the National Science Foundation.

[1] See, for example, Marschak and Radner (1972, ch. 2).

[2] This is true, in particular, of the information measure commonly associated with the names of Shannon and Wiener; see, again, Marschak and Radner (1972, ch. 2).

*Bayesian Models in Economic Theory, edited by M. Boyer and R.E. Kihlstrom*

The net value of a particular information structure may be defined as the maximum expected utility that can be achieved using it (after subtracting the cost). The object of this chapter is to show that, for an important class of decision problems, a small amount of information has a negative marginal net value whenever the marginal cost of information is strictly positive.[3] This result has some important implications. In particular, it implies: (1) if there is some amount of information that has positive net value, then value cannot be a concave function of the amount of information, i.e. there must be *increasing returns to information* over some range of the parameter; (2) the demand for information will not be a continuous function of its price; and (3) in economic activities where information is important, specialization may be common.

In section 2 we present the general theorem and its proof, while the remaining sections develop several applications of the general theorem: to consumer behavior (the optimal allocation of a portfolio); to problems of screening; and to a linear prediction problem.

Before presenting the proof, it may be worthwhile to present briefly the general context of the problem we are investigating. We consider a decision-maker facing a set of decisions. The outcome (payoff) depends on the decision as well as on the state of nature. There is a set of signals. Information provides the decision-maker with a basis for translating the signals into different actions. This basis is quantified in terms of the conditional probabilities of the signals given the states of nature. A completely noninformative information structure is one for which the conditional probability distribution of signals is the same for all states of nature. With a noninformative information structure, the decision-maker might just as well take the same action regardless of the signal received. A "small amount of information" leads to a small change in the conditional probabilities of signals given states. To prove our main result we shall require: (1) in the neighborhood of zero information these conditional probabilities vary smoothly (differentiably), and the corresponding optimal decisions vary continuously, with the amount of information; (2) the optimal decision for zero information is the same for all signals; and (3) utility is a continuous function.

## 2. The basic theorem and some corollaries

We consider a family of decision problems, corresponding to a family of alternative information structures. Let $S$ denote the set of alternative *states of the environment*, and $Y$ the set of alternative *information signals*. $S$ and $Y$ are finite.

An *information structure* is a Markov matrix $((p_{sy}))$, where $p_{sy}$ is the conditional probability of signal $y$ given state $s$. The family of available information structures is indexed by a (real-valued) *parameter* $\theta$. The index set, $\Theta$, is taken to be an interval $[0, \theta_1]$, where $\theta_1$ is strictly positive.

[3] A similar result holds if we define the net value of a particular information structure in terms of its dollar-equivalent value, in those problems in which such a dollar-equivalent is meaningful (see below, end of section 2).

In response to an information signal, the decision-maker chooses an *action*; the set $A$ of alternative actions is a subset of $K$-dimensional Euclidean space, $\mathbb{R}^K$. A *decision function* is a function, say $d = (d_y)$ from $Y$ to $A$, where $d_y$ denotes the action chosen in response to signal $y$.

There is a further *constraint* on action, determined by an inequality

$$g(a, \theta) \leqslant 0, \tag{1}$$

where $g$ is a real-valued function defined on $A \times \Theta$. For each $\theta$ in $\Theta$, let $\mathscr{A}(\theta)$ denote the set of actions in $A$ that satisfy the constraint (1), and let $\mathscr{D}(\theta)$ denote the set of decision functions $d$ that satisfy

$$g(d_y, \theta) \leqslant 0 \quad \text{all} \quad y. \tag{2}$$

Call $\mathscr{D}(\theta)$ the set of admissible decision functions, given $\theta$.

Given the state $s$, the action $a$, and the parameter $\theta$, the *payoff* to the decision-maker is $u_s(a, \theta)$.

Let $\phi_s$ denote the prior probability of state $s$. The *expected value of a decision function* $d$, given $\theta$, is:

$$U(d, \theta) \equiv \sum_{s,y} \phi_s p_{sy}(\theta)\, u_s(d_y, \theta). \tag{3}$$

The value of the information structure $\theta$ is

$$V(\theta) \equiv \sup \{U(d, \theta): \ d \ \text{in} \ \mathscr{D}(\theta)\}. \tag{4}$$

A decision function is *optimal for* $\theta$ if it is admissible for $\theta$ and its expected value achieves the supremum, $V(\theta)$. From (3) it is clear that, if $d$ is a decision function that is optimal for $\theta$, then for every $y$ the decision $d_y$ is an action that maximizes

$$\sum_s \phi_s p_{sy}(\theta)\, u_s(a, \theta)$$

subject to $a$ in $\mathscr{A}(\theta)$.

We assume that among the available information structures there is one that is completely noninformative;[4] this one will be designated by $\theta = 0$. By the definition of "noninformative", $p_{sy}(0)$ is independent of $s$, for each $y$; we denote this common value by $p_y^0$. Thus,

$$p_{sy}(0) = p_y^0 \quad \text{all} \quad y. \tag{5}$$

If $d$ is optimal for $\theta = 0$, then for each $y$ the decision $d_y$ maximizes

[4] For a systematic treatment of information structures and their comparison, see McGuire (1972).

$$\sum_s \phi_s p_{sy}(0)\, u_s(a, 0)$$

on $\mathscr{A}(\theta)$; by the definition of noninformative information structure, (5), this last expression is equal to:

$$p_y^0 \sum_s \phi_s u_s(a, 0).$$

Hence, for $\theta = 0$, we may take the optimal action to be independent of the signal $y$, say $a^0$; i.e.

$$d_y = a^0 \quad \text{for all} \quad y. \tag{6}$$

We shall call a decision function *flat* if it results in the same action for all information signals $y$. In this terminology, we have just shown that for $\theta = 0$ there is an optimal decision function that is flat.

In the interpretation of the following theorem, the reader may have in mind the situation in which larger values of $\theta$ correspond to information structures that are more informative (in some sense), but also more "costly". The increased "cost" may be reflected in a lower payoff ($u_s$ a decreasing function of $\theta$), or a smaller constraint set ($g$ an increasing function of $\theta$), or both. However, for a strict interpretation of the theorem, there is no need to suppose that larger $\theta$ corresponds to more informative structures, but only that $\theta = 0$ corresponds to a noninformative structure.

We shall be concerned with a family of decision functions (one for each $\theta$), which we can represent as a mapping $D$ from $\Theta$ to the set of decision functions; thus, for every $\theta$, $D(\theta)$ is an admissible decision function for $\theta$. Since the set $Y$ of signals is finite, each decision function may be thought of as a point in a Euclidean vector space of dimension equal to $K$ times the number of signals in $Y$. Thus, a family $D$ of decision functions is itself a function from $\Theta$ to this vector space.

**Theorem.** If (i) for every $s$ in $S$ and $y$ in $Y$, $p_{sy}(\cdot)$ is differentiable at $\theta = 0$, and (ii) for every $s$ in $S$ and $a$ in $A$, $u_s(a, \cdot)$ is monotone nonincreasing and $g(a, \cdot)$ is monotone nondecreasing on $\Theta$, and $u_s(\cdot, \cdot)$ is continuous on $A \times \Theta$, then for any family $D$ of decision functions that is both flat and continuous at $\theta = 0$:

$$\limsup_{\theta \to 0} \frac{U[D(\theta), \theta] - U[D(0), 0]}{\theta} \leqslant 0.$$

*Proof.* From (3):

$$U[D(\theta), \theta] - U[D(0), 0] = T_1(\theta) + T_2(\theta),$$

where

$$T_1(\theta) = \sum_{s,y} \phi_s p_{sy}(\theta)\, u_s[D_y(\theta), \theta] - \sum_{s,y} \phi_s p_{sy}(0)\, u_s[D_y(\theta), \theta],$$

$$T_2(\theta) = \sum_{s,y} \phi_s p_{sy}(0)\, u_s[D_y(\theta), \theta] - \sum_{s,y} \phi_s p_{sy}(0)\, u_s[D_y(0), 0].$$

Since $u_s(a, \cdot)$ is monotone nonincreasing on $\Theta$,

$$u_s[D_y(\theta), \theta] \leqslant u_s[D_y(\theta), 0]. \tag{7}$$

Since $g(a, \cdot)$ is monotone nondecreasing, the decision function $D(\theta)$ is admissible for $\theta = 0$; but $D(0)$ is optimal for $\theta = 0$, so that

$$\sum_{s,y} \phi_s p_{sy}(0)\, u_s[D_y(\theta), 0] \leqslant \sum_{s,y} \phi_s p_{sy}(0)\, u_s[D_y(0), 0]. \tag{8}$$

Combining (7) and (8), we get:

$$T_2(\theta) \leqslant 0,$$

so that

$$\limsup_{\theta \to 0} \frac{T_2(\theta)}{\theta} \leqslant 0. \tag{9}$$

We may rewrite $T_1(\theta)$ as

$$T_1(\theta) = \sum_{s,y} \phi_s [p_{sy}(\theta) - p_{sy}(0)]\, u_s[D_y(\theta), \theta].$$

Since $D(0)$ is flat, there is an action, say $a^0$, such that $D_y(0) = a^0$ for all signals $y$. Also, $D$ is continuous at 0, and $u_s(\cdot, \cdot)$ is continuous for each $s$. Therefore

$$\lim_{\theta \to 0} u_s[D_y(\theta), \theta] = u_s(a^0, 0),$$

and, since $p_{sy}(\cdot)$ is differentiable at 0,

$$\begin{aligned} \lim_{\theta \to 0} \frac{T_1(\theta)}{\theta} &= \sum_{s,y} \phi_s p'_{sy}(0)\, u_s(a^0, 0) \\ &= \sum_s \phi_s u_s(a^0, 0) \sum_y p'_{sy}(0). \end{aligned}$$

For every $s$ and $\theta$:

$$\sum_y p_{sy}(\theta) = 1;$$

hence,

$$\begin{aligned} &\sum_y p'_{sy}(0) = 0, \\ &\lim_{\theta \to 0} \frac{T_1(\theta)}{\theta} = 0. \end{aligned} \tag{10}$$

Therefore

$$\limsup_{\theta \to 0} \frac{T_1(\theta) + T_2(\theta)}{\theta} \leqslant 0,$$

which completes the proof.

A family $D$ of decision functions is *optimal* if, for every $\theta$, $D(\theta)$ is optimal for $\theta$. The following corollary follows immediately from the theorem.

**Corollary.** If hypotheses (i) and (ii) of the theorem are satisfied, and if there is an optimal family of decision functions that is both flat and continuous at $\theta = 0$, then

$$\limsup_{\theta \to 0} \frac{V(\theta) - V(0)}{\theta} \leqslant 0. \tag{11}$$

To see that if the marginal cost of information is *strictly* positive, then the marginal net value of information (at $\theta = 0$) will be *strictly* negative, we consider a suitably specialized formulation. Suppose that: (i) the "gross" outcome of action $a$ in state $s$ is a real number (e.g. money), denoted by $w_s(a)$; (ii) the cost of information structure $\theta$ is $C(\theta)$; and (iii) the "utility" of the net outcome is $v_s[w_s(a) - C(\theta)]$. Thus, the payoff function $u_s(\cdot, \cdot)$ is given by:

$$u_s(a, \theta) = v_s[w_s(a) - C(\theta)]. \tag{12}$$

Assume that $v_s$ and $C$ have continuous and strictly positive derivatives, that $C(0) = 0$, and that $w_s$ is continuous. Consider the function $T_2$ defined in the proof of the theorem. We may rewrite this function as:

$$\begin{aligned} T_2(\theta) = \sum_{s,y} \phi_s p_{sy}(0) \{ & u_s[D_y(\theta), \theta] - u_s[D_y(\theta), 0] \\ & + u_s[D_y(\theta), 0] - u_s[D_y(0), 0]\}. \end{aligned} \tag{13}$$

Concentrating first on the first difference in the curly brackets of (13), and making use of the special formulation (12), we have:

$$\begin{aligned} u_s[D_y(\theta), \theta] - u_s[D_y(\theta), 0] &= v_s[w_s(D_y[\theta]) - C(\theta)] - v_s[w_s(D[\theta])] \\ &= -v_s'[w_s(D_y[\theta])]\, C(\theta) + 0(\theta), \end{aligned} \tag{14}$$

where $0(\theta)/\theta$ converges to 0 as $\theta$ approaches 0. Hence,

$$\lim_{\theta \to 0} \frac{u_s[D_y(\theta), \theta] - u_s[D_y(\theta), 0]}{\theta} = -v_s'[w_s(D_y[0])]\, C'(0), \tag{15}$$

which is strictly negative. Combining (15) and (8) with (13) gives:

$$\limsup_{\theta \to 0} \frac{T_2(\theta)}{\theta} < 0,$$

which is a strengthening of (9). Hence, we can strengthen the conclusion of the theorem to:

$$\limsup_{\theta \to 0} \frac{U[D(\theta), \theta] - U[D(0), 0]}{\theta} < 0. \tag{16}$$

A corresponding analysis could be made to cover the case in which the effect of the cost of information is to restrict the set of admissible decisions, i.e. the case in which the constraint (2) is binding at $\theta = 0$ and the function $g(a, \cdot)$ is strictly increasing in $\theta$, for every action $a$.

## 3. A linear prediction problem

Our first example does not quite fit our theoretical framework in that the set $S$ of states of the environment and the set $Y$ of signals are both infinite. Nevertheless, the example provides a brief illustration of the basic result in a framework that is probably quite familiar to most readers.

Assume that $y$ and $s$ are normally distributed random variables with

$$\text{Corr}\,(y, s) = \rho. \tag{17}$$

We choose our units so that

$$\text{Var}\,(s) = \text{Var}\,(y) = 1, \tag{18}$$

and we choose our origins so that

$$\text{E}s = \text{E}y = 0. \tag{19}$$

Let $\theta$, the information structure, be represented by $\rho$, the correlation coefficient between $s$ and $y$:

$$\theta \equiv \rho. \tag{20}$$

Let the gross payoff associated with taking some action $a$ if state $s$ occurs be:

$$-(a - s)^2,$$

and the net payoff be:

$$u(s, a, \theta) = -(a - s)^2 - C(\theta), \tag{21}$$

where $C(\theta)$ is the cost of obtaining the information structure $\theta$. Assume that

$$C(0) = 0 \quad \text{and} \quad C'(0) > 0. \tag{22}$$

Thus, observing a variable that is uncorrelated with $s$ is costless; observing any variable that is correlated with $s$ has some cost.

The optimal decision rule, given $\theta$, is the regression of $s$ on $y$:

$$\begin{aligned} D_y(\theta) &= \mathrm{E}(s \mid y) \\ &= \rho y. \end{aligned} \tag{23}$$

Substituting (23) into (21), we immediately obtain the *value* of the information structure $\theta$, i.e. the expected payoff using the optimal decision rule:

$$V(\theta) = \theta^2 - 1 - C(\theta), \tag{24}$$

so that

$$V'(\theta) = 2\theta - C'(\theta), \tag{25}$$

$$V'(0) = -C'(0) < 0. \tag{26}$$

If, instead of (20), we represent our information structure by

$$\theta = \rho^2, \tag{20'}$$

then, instead of (24), we obtain:

$$V(\theta) = \theta - 1 - \tilde{C}(\theta), \tag{24'}$$

where $\tilde{C}(\theta) \equiv C(\theta^{1/2})$. Thus,

$$V'(0) = 1 - \tilde{C}'(0), \tag{25'}$$

which could be positive. This appears to contradict our theorem. However, with this second representation of the "quantity" of information, condition (i) of the basic theorem, is violated. The conditional distribution of $y$ given $s$ is normal with mean $\rho s$ and variance $(1 - \rho^2)$; the corresponding conditional density is

$$p_{sy} = \frac{1}{[2\pi(1-\rho^2)]^{1/2}} \exp\left(-\frac{(y-\rho s)^2}{2(1-\rho^2)}\right).$$

One can verify that if $\theta = \rho^2$, then

$$\lim_{\theta \to 0} \frac{\mathrm{d}p_{sy}}{\mathrm{d}\theta} = +\infty,$$

so that $p_{sy}$ would not be differentiable in $\theta$ at $\theta = 0$.

## 4. A portfolio model[5]

There are $n$ securities, each of which has a price of one dollar, and $n$ corresponding states. Security $s$ pays $r_s$ if state $s$ occurs, and nothing otherwise. The investor has an initial wealth $W$, of which he allocates part to the purchase of information and the remainder to the purchase of securities. Let $C(\theta)$ denote the cost of information structure $\theta$, and let

$$a_{ys}(\theta)\,[W - C(\theta)]$$

denote the remaining wealth allocated to the purchase of security $s$ if signal $y$ is observed; then the allocation proportions, $a_{ys}(\theta)$, must satisfy

$$\sum_s a_{ys}(\theta) = 1 \quad \text{all} \quad y. \tag{27}$$

If signal $y$ is observed and state $s$ occurs, the payoff (return) from the purchase of securities will be:

$$r_s a_{ys}(\theta)\,[W - C(\theta)].$$

Assume that the investor's utility function is logarithmic. Then the expected utility associated with an information structure $\theta$ and allocation proportions $a_{ys}(\theta)$ is:

$$\sum_{s,y} \phi_s p_{sy}(\theta) \log \{r_s a_{ys}(\theta)\,[W - C(\theta)]\}, \tag{28}$$

where, as in section 2, $\phi_s$ is the prior probability of state $s$, and $p_{sy}(\theta)$ is the conditional probability of signal $y$ given state $s$ (for the information structure $\theta$). The (unconditional) probability that signal $y$ is observed is:

$$q_y(\theta) \equiv \sum \phi_s p_{sy}(\theta),$$

[5] Models like this one have been discussed frequently. See, for example, Arrow (1972) and the references given there.

and the conditional ("posterior") probability of state $s$, given signal $y$, is:

$$\Pi_{ys}(\theta) \equiv \frac{\phi_s p_{sy}(\theta)}{q_y(\theta)} .$$

As in section 2, $\theta = 0$ denotes a noninformative information structure, so that

$$\begin{aligned} &p_{sy}(0) = p_y^0 \quad \text{all} \quad s, \\ &q_y(0) = p_y^0, \quad \Pi_{ys}(0) = \phi_s. \end{aligned} \tag{29}$$

With the foregoing notation, we can express the expected utility (28) as:

$$\sum_y q_y(\theta) \sum_s \Pi_{ys}(\theta) \log \{r_s a_{ys}(\theta) [W - C(\theta)]\}. \tag{30}$$

It is straightforward to verify that the optimal allocation proportions are equal to the corresponding conditional probabilities, i.e.

$$\hat{a}_{ys}(\theta) = \Pi_{ys}(\theta).$$

Therefore the net value of the information structure $\theta$ is:

$$V(\theta) = \sum_y q_y(\theta) \sum_s \Pi_{ys}(\theta) \log \{r_s \Pi_{sy}(\theta) [W - C(\theta)]\}. \tag{31}$$

The expression (31) for the value of $\theta$ can be rewritten in a way that relates it to the Shannon–Wiener measure of information:

$$V(\theta) = I(\theta) + f(\theta) + \log \{W - C(\theta)] + \sum_s \phi_s \log \phi_s, \tag{32}$$

where

$$I(\theta) = \sum_y q_y(\theta) \sum_s \Pi_{ys}(\theta) \log \Pi_{ys}(\theta) - \sum_s \phi_s \log \phi_s, \tag{33}$$

$$f = \sum_s \phi_s \log r_s. \tag{34}$$

Recall that the Shannon–Wiener measure of "uncertainty" in a probability distribution $P = (P_i)$ is defined as:

$$-\sum_i P_i \log P_i.$$

If signal $y$ is observed, the conditional uncertainty about $s$ is:

$$-\sum_s \Pi_{ys}(\theta) \log \Pi_{ys}(\theta).$$

Therefore $I(\theta)$ is the expected *reduction* in uncertainty about $s$ associated with the information structure $\theta$.

For example, consider the special case in which there are two states and two signals, with

$$p_{11} = p_{22} = 1/2 + \theta, \quad p_{12} = p_{21} = 1/2 - \theta,$$
$$\phi_1 = \phi_2 = 1/2, \quad 0 \leqslant \theta \leqslant 1/2.$$

Then

$$\Pi_{11}(\theta) = \Pi_{22}(\theta) = 1/2 + \theta, \quad \Pi_{12}(\theta) = \Pi_{21}(\theta) = 1/2 - \theta,$$
$$I(\theta) = (1/2 + \theta) \log (1/2 + \theta) + (1/2 - \theta) \log (1/2 - \theta) + \log 2.$$

Here $I(\theta)$ varies from 0 ($\theta = 0$) to log 2 ($\theta = 1/2$), the latter corresponding to "perfect information".

We return now to our more general formula (32), and calculate the marginal net value of information. First, we write $I(\theta)$ as

$$I(\theta) = \sum_{s,y} \phi_s p_{sy}(\theta) \log \Pi_{ys}(\theta) - \sum_s \phi_s \log \phi_s.$$

Differentiation with respect to $\theta$ yields:

$$\begin{aligned} I'(\theta) &= \sum_{s,y} \left[ \frac{\phi_s p_{sy}(\theta) \Pi'_{ys}(\theta)}{\Pi_{sy}(\theta)} + \phi_s p'_{sy}(\theta) \log \Pi_{ys}(\theta) \right] \\ &= \sum_{s,y} [q_y(\theta) \Pi'_{ys}(\theta) + \phi_s p'_{sy}(\theta) \log \Pi_{ys}(\theta)] \\ &= \sum_y q_y(\theta) \sum_s \Pi'_{ys}(\theta) + \sum_s \phi_s \sum_y p'_{sy}(\theta) \log \Pi_{ys}(\theta). \end{aligned} \tag{35}$$

Since $\Sigma_s \Pi_{ys}(\theta) = 1$ for all $y$ and $\theta$, $\Sigma_s \Pi'_{ys}(\theta) = 0$, all $s$ and $\theta$. Similarly

$$\sum_y p'_{sy}(\theta) = 0, \quad \text{all} \quad s \quad \text{and} \quad \theta. \tag{36}$$

Therefore, letting $\theta = 0$, and using (29):

$$\begin{aligned} I'(0) &= 0 + \sum_s \phi_s \log \phi_s \sum_y p'_{sy}(0) \\ &= 0. \end{aligned} \tag{37}$$

Since $f$ is independent of $\theta$, and $C(0) = 0$,

$$V'(0) = -\frac{C'(0)}{W}. \tag{38}$$

Thus, if the marginal cost of information is strictly positive, the marginal net value of information is negative at $\theta = 0$.

In the above analysis we calculated the value of information in terms of "utility units"; alternatively, we could have calculated the amount $\hat{V}(\theta)$ the investor would have been willing to give up to acquire the given information structure, i.e. $\hat{V}(\theta)$ is given by the solution of:

$$I(\theta) + f + \log [W - \hat{V}(\theta)] = I(0) + f + \log W,$$

or equivalently,

$$-\log \frac{W - \hat{V}(\theta)}{W} = [I(\theta) - I(0)]. \tag{39}$$

Differentiating (39) with respect to $\theta$ and solving for $\hat{V}'(\theta)$ one gets:

$$\hat{V}'(\theta) = [W - \hat{V}(\theta)]\, I'(\theta) \geqslant 0,$$

$$\hat{V}'(0) = [W - \hat{V}(0)]\, I'(0) = 0.$$

Thus, although the marginal "money value" of information is non-negative, it is zero at $\theta = 0$.

## 5. A general screening model[6]

A population consists of $I$ types of individual, and each individual is to be assigned by a firm to one of $J$ jobs. Let $N_i$ denote the number of individuals in the population who are of type $i$ (all $N_i > 0$). An examination $\theta$ is administered to each individual, and results in a "label" $y$ for that individual ($y = 1, ..., I$); this label is the signal on the basis of which the individual is assigned to a job $j$. The decision rule, $(d_y)$, relates the label of the individual to the job to which he is assigned. We represent this by the matrix $(d_{yj})$, where

$$d_{yj} = \begin{cases} 1 & \text{if an individual of label } y \text{ is assigned to job } j, \text{ and} \\ 0 & \text{otherwise.} \end{cases} \tag{40}$$

If an individual of type $i$ is assigned to job $j$, the dollar value of his output is $b_{ij}$. Let $p_{iy}(\theta)$ denote the probability that an individual receives the label $y$, conditional on being of type $i$. Then, for a given decision rule, $(d_{yj})$, the expected value of output is:

[6] For a more extensive discussion of screening models, see Stiglitz (1974).

$$Q(\theta) \equiv \sum_i N_i \sum_y p_{iy}(\theta) \sum_j d_{yj} b_{ij}. \tag{41}$$

If the cost of the examination $\theta$ is $C(\theta)$, then the expected net profit to the firm is

$$Q(\theta) - C(\theta), \tag{42}$$

and the firm's objective is to choose the examination $\theta$ and the decision rule $(d_{yj})$ so as to maximize this expected net profit.

Let $G(\theta)$ denote the maximum expected value of output, given $\theta$, i.e. the maximum of (41) with respect to the decision rule $(d_{yj})$. If we write (41) as

$$Q(\theta) = \sum_y \sum_j d_{yj} \sum_i N_i p_{iy}(\theta) b_{ij}, \tag{43}$$

it is easy to see that this maximum is attained by assigning each individual with label $y$ to a job $j$ for which

$$\sum_i N_i p_{iy}(\theta) b_{ij}$$

is maximum. Hence,

$$G(\theta) = \sum_y \max_j \sum_i N_i p_{iy}(\theta) b_{ij}, \tag{44}$$

and the (expected) value of the examination $\theta$ is:

$$V(\theta) = G(\theta) - C(\theta). \tag{45}$$

As we have formulated the screening problem, the decision variables are discrete, rather than continuous. However, if we allow "randomized" decisions, then we can reinterpret $d_{yj}$ as the probability that a person with label $y$ is assigned to job $j$. With this interpretation, (40) is replaced by:

$$\begin{cases} d_{yj} \geqslant 0 & \text{all } y \text{ and } j, \\ \sum_j d_{yj} = 1 & \text{all } y. \end{cases} \tag{46}$$

Because of the linearity of $Q(\theta)$ in the variables $(d_{yj})$, any decision rule that was optimal for the first formulation is optimal for the second, and the formulae (44) and (45) are still valid for the value of the examination.

We now examine the conditions under which $G'(0) = 0$. There are two cases, according as the optimal assignment for $\theta = 0$ is or is not unique. Recall that for a noninformative examination ($\theta = 0$):

$$p_{iy}(0) = p_y^0 \quad \text{all } i. \tag{47}$$

Hence, for $\theta = 0$:

$$G(0) = \sum_y p_y^0 \max_j \sum_i N_i b_{ij}, \tag{48}$$

and each individual, regardless of label, should be assigned to a job $j^0$ for which $\Sigma_i N_i b_{ij}$ is at a maximum. (Of course, if $p_y^0 = 0$ for some label $y$, then all assignments for that label are optimal!) There are two cases.

*Case I. $j^0$ is unique, and $p_y^0 > 0$ for all $y$.*

In this case,

$$\sum_i N_i b_{ij}0 > \sum_i N_i b_{ij} \quad \text{all} \quad j \neq j^0. \tag{49}$$

If $p_{ij}(\theta)$ is continuous in $\theta$, then for all sufficiently small $\theta$, say $0 \leqslant \theta \leqslant \theta_0$, and all $y$:

$$\sum_i N_i p_{iy}(\theta)\, b_{ij}\, 0 > \sum_i N_i p_{iy}(\theta)\, b_{ij}, \quad j \neq j^0, \tag{50}$$

and so

$$G(\theta) = G(0), \quad 0 \leqslant \theta \leqslant \theta^0. \tag{51}$$

Hence,

$$G'(0) = 0, \tag{52}$$

and if $C'(0) > 0$, $V'(0) < 0$, so that *the net marginal value of the examination is negative for $\theta$ near zero.*

*Case II. $j^0$ is not unique, and/or $p_y^0 = 0$ for some $y$.*

To illustrate the implications of this case, we consider the special case of two types and two jobs ($I = J = 2$), with $p_1^0$ and $p_2^0$ both strictly positive. If $j^0$ is not unique, then

$$N_1 b_{11} + N_2 b_{21} = N_1 b_{12} + N_2 b_{22}. \tag{53}$$

For $\theta$ small but not zero, it will typically be true that, for both $y$:

$$N_1 p_{1y}(\theta)\, b_{11} + N_2 p_{2y}(\theta)\, b_{21} \neq N_1 p_{1y}(\theta)\, b_{12} + N_2 p_{2y}(\theta)\, b_{22}. \tag{54}$$

Without further loss of generality, we may suppose that the left-hand side of (54) is larger than the right-hand side for $y = 1$, and therefore the right-hand side is the larger for $y = 2$ (this last follows from (53) and the fact that

$$p_{i1}(\theta) + p_{i2}(\theta) = 1 \tag{55}$$

for each $i$). In other words, for $\theta$ small but not zero, individuals with label 1 should be assigned to job 1, and those with label 2 should be assigned to job 2. It follows that

$$\begin{aligned} G(\theta) &= [N_1 p_{11}(\theta) b_{11} + N_2 p_{21}(\theta) b_{21}] + [N_1 p_{12}(\theta) b_{12} + N_2 p_{22}(\theta) b_{22}] \\ &= N_1 [p_{11}(\theta) b_{11} + p_{12}(\theta) b_{12}] + N_2 [p_{21}(\theta) b_{21} + p_{22}(\theta) b_{22}], \end{aligned}$$

and from this and (55) that

$$G'(\theta) = N_1 p'_{11}(\theta)(b_{11} - b_{12}) + N_2 p'_{22}(\theta)(b_{22} - b_{21}). \tag{56}$$

But, from (53):

$$N_1 (b_{11} - b_{12}) = N_2 (b_{22} - b_{21}), \tag{57}$$

so

$$G'(\theta) = [p'_{11}(\theta) + p'_{22}(\theta)] N_1 (b_{11} - b_{12}). \tag{58}$$

We now show that $G'(0) > 0$ if $p'_{11}(0)$ and $p'_{22}(0)$ are not both zero, i.e. if the examination is informative for small $\theta$. First we note that, by our convention that (for small $\theta$) individuals with label 1 should be assigned to job 1:

$$\begin{aligned} &N_1 p_{11}(\theta) b_{11} + N_2 p_{21}(\theta) b_{21} > N_1 p_{11}(\theta) b_{12} + N_2 p_{21}(\theta) b_{22}, \\ &p_{11}(\theta) N_1 (b_{11} - b_{12}) > p_{21}(\theta) N_2 (b_{22} - b_{21}), \end{aligned}$$

so, using (53) again, and the fact that $N_1 > 0$:

$$[p_{11}(\theta) - p_{21}(\theta)][b_{11} - b_{12}] > 0. \tag{59}$$

Thus $[p_{11}(\theta) - p_{21}(\theta)]$ and $[b_{11} - b_{12}]$ are of the same sign. This has a natural interpretation. The conditional probabilities that an individual is of type 1 and type 2, respectively, given the label 1, are proportional to $p_{11}(\theta)$ and $p_{21}(\theta)$. If, conditional on the label 1, an individual is more likely to be of type 1 than of type 2 $[p_{11}(\theta) > p_{21}(\theta)]$, and if an individual of type 1 is more productive in job 1 than in job 2 $(b_{11} > b_{12})$, then an individual with label 1 should be assigned to job 1. The same assignment should be made if the individual is more likely to be of type 2 than of type 1 $[p_{11}(\theta) < p_{21}(\theta)]$, and is more productive in job 2 than in job 1 $(b_{11} < b_{12})$.

Suppose that $b_{11} > b_{12}$; then $p_{11}(\theta) > p_{21}(\theta)$, or

$$p_{11}(\theta) + p_{22}(\theta) - 1 > 0 \tag{60}$$

for $\theta$ small but $> 0$. However,

$$p_{11}(0) + p_{22}(0) - 1 = p_1^0 + p_2^0 - 1 = 0, \tag{61}$$

so that, comparing (60) with (61):

$$p'_{11}(0) + p'_{22}(0) \geqslant 0. \tag{62}$$

Therefore, from (58):

$$G'(0) \geqslant 0, \quad \text{and}$$
$$G'(0) > 0 \quad \text{if} \quad p'_{11}(0) \quad \text{and} \quad p'_{22}(0) \quad \text{are not both zero.} \tag{63}$$

The same conclusion follows, by similar reasoning, if $b_{11} < b_{12}$.

In summary, for case II, *the marginal net value of the examination at* $\theta = 0$ *will be positive if*: (1) $p'_{11}(0)$ *and* $p'_{22}(0)$ *are both zero, and* (2) *the marginal cost of the examination at* $\theta = 0$ *is not too large* $[C'(0) < G'(0)]$.

The situation just described in case II contradicts the conclusion of our main theorem. However, in this situation a family of optimal decision rules cannot be both flat and continuous at $\theta = 0$. If it is to be flat at $\theta = 0$, then it must assign all individuals to the same job. On the other hand, if either $p'_{11}(0)$ or $p'_{22}(0)$ are different from 0, then individuals with different labels should be assigned different jobs; so that if a family of optimal decision rules were flat at $\theta = 0$ it would be discontinuous there.

### *A reformulation*

We can easily reformulate our screening problem to conform to the postulates of our theorem. Assume there is a continuum of possible types of jobs. An individual of type $i$ assigned to job $a$ has a productivity of

$$K - (a - \alpha_i)^2. \tag{64}$$

Thus, there is a social loss from misassigning individuals, which is an increasing function of the magnitude of the misassignment. It is easy to show that an optimal decision rule with an examination of accuracy $\theta$ must satisfy

$$D_y(\theta) = \frac{\sum_i N_i p_{iy}(\theta)\, \alpha_i}{\sum_i N_i p_{iy}(\theta)} \equiv \bar{\alpha}_y(\theta), \tag{65}$$

for all labels $y$ with $\Sigma_i N_i p_{iy}(\theta) > 0$, where $D_y(\theta)$ is the optimal assignment of an individual with label $y$. Thus (given $\theta$), the maximum expected value of output is:

$$\begin{aligned} G(\theta) &= \sum_{i,y} N_i p_{iy}(\theta)\,[K - (\overline{\alpha}_y(\theta) - \alpha_i)^2] \\ &= NK - \sum_{i,y} N_i p_{iy}(\theta)\,(\alpha_i^2 - \overline{\alpha}_y(\theta)^2). \end{aligned} \tag{66}$$

For $\theta = 0$, an optimal decision rule must satisfy:

$$D_y(0) = \frac{\sum_i N_i \alpha_i}{\sum_i N_i} \equiv \overline{\alpha} \tag{67}$$

for all labels $y$ with $p_y^0 > 0$. In particular, there is a unique flat optimal decision rule for $\theta = 0$, namely $D_y(0) = \overline{\alpha}$ for all $y$. On the other hand, if we let $\theta$ tend to 0 in (65) we get:

$$\lim_{\theta \to 0} D_y(\theta) = \begin{cases} \overline{\alpha} & \text{if } \; p_y^0 > 0, \\ \dfrac{\sum_i N_i p'_{iy}(0)\,\alpha_i}{\sum_i N_i p'_{iy}(0)} & \text{if } \; p_y^0 = 0, \end{cases} \tag{68}$$

for all labels $y$ with $\Sigma_i N_i p_{iy}(\theta) > 0$ for $\theta$ near zero, provided the second line on the right-hand side of (68) is well defined. (The second line follows from l'Hôpital's Rule.) Comparing (67) and (68), we see that a family of optimal decision rules that is continuous at $\theta = 0$ need not be flat at $\theta = 0$, unless $p_y^0 > 0$ for every label $y$. The expression $\Sigma_i N_i p_{iy}(\theta)$ is the expected number of individuals who receive the label $y$ (given the examination $\theta$), and

$$\lim_{\theta \to 0} \sum_i N_i p_{iy}(0) = N p_y^0. \tag{69}$$

Thus, it is possible for the marginal net value of the examination to be positive only if there exists a label that does not appear ($p_y^0 = 0$) with the uninformative examination ($\theta = 0$), but does appear with positive frequency ($\Sigma_i N_i p_{iy}(\theta) > 0$) with an examination that is slightly informative ($\theta > 0$ but small). (This last is, of course, a necessary but not sufficient condition for $G'(0)$ to be strictly positive.)

Consider, for example, the special case of two types, with

$$\begin{aligned} N_1 &= N_2, \\ p_{11}(\theta) &= p_{22}(\theta) = 1/2 + \theta, \\ p_{12}(\theta) &= p_{21}(\theta) = 1/2 - \theta, \quad 0 \leqslant \theta \leqslant 1/2. \end{aligned} \tag{70}$$

Then the (unique) family of optimal decision rules is:

$$D_1(\theta) = \bar{\alpha} + \theta(\alpha_1 - \alpha_2),$$
$$D_2(\theta) = \bar{\alpha} + \theta(\alpha_2 - \alpha_1), \tag{71}$$
$$\bar{\alpha} \equiv \frac{\alpha_1 + \alpha_2}{2},$$

and the maximum expected value of output, given $\theta$, is:

$$G(\theta) = NK - \left(\frac{N}{2}\right)(\alpha_1^2 + \alpha_2^2) + \left(\frac{N}{2}\right)(\bar{\alpha}_1^2 + \bar{\alpha}_2^2). \tag{72}$$

Hence, $G'(0) = 0$.

Similar considerations show that for "hierarchical" screening (as opposed to "job-matching" screening of the kind described above), if screening is undertaken at all it will be undertaken above some minimum level of quality (informativeness). We consider here only the special case just discussed, with two groups in the population. Now, however, productivity does not depend on job assignment. Rather, type 1 individuals are always more productive than type 2 individuals. Let $\alpha_1$ and $\alpha_2$ be their respective productivities. Assume $\alpha_1 > \alpha_2$. Each type of individual knows what type he is, but firms do not. Type 1 individuals are considering whether to attempt to persuade the government to subject all individuals to a screening system. (Assume, for instance, that there is one more type 1 individual than type 2.) If they get labeled $i$, they receive the mean marginal productivity of individuals labeled $i$. Thus, their expected income under a screening system of accuracy $\theta$ is:

$$[\bar{\alpha} + \theta(\alpha_1 - \alpha_2)](1/2 + \theta) + [\bar{\alpha} - \theta(\alpha_1 - \alpha_2)](1/2 - \theta),$$

so the increment in expected income is:

$$2\theta^2(\alpha_1 - \alpha_2).$$

It is immediately apparent that, even if the total costs of screening are borne uniformly over the population, there will be some minimal quality of screening before the more able wish to have it undertaken, i.e. letting

$$V(\theta) = \theta^2(\alpha_1 - \alpha_2) - C(\theta), \quad V'(0) = -C'(0) < 0.$$

The question naturally arises: How can we relate this result, which does not appear to arise from an optimization problem but rather from a description of market equilibrium, to our theorem?

Consider the problem of maximizing the expected income of the first group, subject to the constraints that

$$a_1(1/2+\theta)+a_2(1/2-\theta)<\overline{\alpha}+2\theta^2(\alpha_1-\alpha_2)$$

and

$$a_1+a_2\leqslant\overline{\alpha},$$

or

$$g(a,\theta)\equiv\theta(2a_1-\overline{\alpha})-2\theta^2(\alpha_1-\alpha_2)\leqslant 0, \quad (73)$$

where $a_y$, the action, is the wage paid to a person with label $y$. The interpretation of the constraint is that the mean income of a person of type 1 must be less than or equal to his mean income were each label paid its marginal productivity. The constraint has the property that

$$g_\theta(a,\theta)=(2a_1-\overline{a})-4\theta(\alpha_1-\alpha_2),$$

which is zero at $a_1=a_2$ and $\theta=0$.

Thus, we maximize

$$a_1(1/2+\theta)+a_2(1/2-\theta)-C(\theta) \quad (74)$$

subject to (73).

The optimal decision rule is:

$$D_1(\theta)=\overline{\alpha}+\frac{\theta(\alpha_1-\alpha_2)}{2}, \quad D_2(\theta)=\overline{\alpha}-\frac{\theta(\alpha_1-\alpha_2)}{2}.$$

It is clear that with this formulation the theorem is applicable to our problem.

Not every screening problem has a natural formulation which generates a decision function which is continuous and flat at $\theta=0$. In particular, consider the problem of the optimal extraction of oil. There is some oil with zero extraction costs, and some with positive extraction costs of $e$. The stock of oil is to be consumed this period or next. There is a cost of ascertaining whether any particular well is a high or low extraction-cost well, $C(\theta)$, which is related to the quality of screening. With information structure $\theta$, the probability that oil of type $i$ is labelled $j$ is as shown in table 3.1 below,

Table 3.1

| | Label | |
|---|---|---|
| Type | 1 | 2 |
| 1 | $\lambda(1-\theta)+\theta$ | $(1-\lambda)(1-\theta)$ |
| 2 | $\lambda(1-\theta)$ | $(1-\lambda)(1-\theta)+\theta$ |

where $\lambda$ is the fraction of wells that are of type 1 (each well produces one unit of oil).

It is well known that it is optimal (with a positive interest rate) to extract the low extraction-cost oil first. Assume $\lambda$ is sufficiently large that for all $\theta$ only oil of label 1 is extracted the first period; then the net value of resource savings for better $\theta$ is given by:

$$\frac{\partial V}{\partial \theta} = \frac{rQ_1}{1+r} e(1-\lambda) - C',$$

where $Q_1$ is the consumption of oil in the first period. It is clear that $\partial V/\partial \theta$ can be positive.

## References

Arrow, K.J. (1972), "The Value of and Demand for Information", in: C.B. McGuire and R. Radner, eds., *Decision and Organization* (Amsterdam: North-Holland) ch. 6.

Marschak, J. and R. Radner (1972), *Economic Theory and Teams* (New Haven: Yale University Press).

McGuire, C.B. (1972), "Comparisons of Information Structures", in: C.B. McGuire and R. Radner, eds., *Decision and Organization* (Amsterdam: North-Holland) ch. 5.

Stiglitz, J.E. (1974), "Information and Economic Analysis", paper presented to the AVTE Conference, Manchester (unpublished).