ELICITING PRODUCTION POSSIBILITIES

FROM A WELL-INFORMED MANAGER*

by

William Thomson

Discussion Paper No. 78-96, March 1978

* I have benefited from discussions with James Jordan and Marcel Richter

Center for Economic Research

Department of Economics

University of Minnesota

Minneapolis, Minnesota 55455

ABSTRACT

Eliciting Production Possibilities

From a Well-Informed Manager.

by William Thomson

The cumulative probability distribution F of a random-variable y, the output of a firm, is known to the manager of the firm only. A planning authority wants to elicit from the manager what output can be produced with probability r. The manager is therefore asked the solution of the equation $F(x) = r$. Production takes place and the value of output is observed by both agents. A bonus is then paid to the manager, computed as a function of his answer and of the level of output. We characterize here all the bonus schemes for which announcing the correct solution of $F(x) = \alpha$ is the manager's best strategy. Some important additional properties of these schemes are studied in a second part of the paper.

Introduction

A series of recent papers have been concerned with the issue of incentives in the context of the Soviet firm. Their general structure is the following: information is exchanged between the manager of a given firm and the planning authority or center. Then production takes place and a bonus is paid to the manager on the basis of the variables he and the center announced as well as of the observed output. A paper by Fan [2] proposes a scheme linear in two variables, a target level chosen by the manager, and the final output. It is such that it is optimal for the manager to choose as target the level of output that can be produced with probability $r = 1/2$.

In a second contribution by Bonin [1], a more general scheme is developed for which it is the manager's best strategy to announce the level of output that can be produced with probability r, where r is now any number between 0 and 1, chosen by the center.

Finally, Weitzman [5] discusses the following bonus scheme: a tentative target and a tentative bonus are chosen for a given firm by the central authority. The manager proposes a new target, and a new bonus is computed. Finally, production takes place and the bonus is adjusted again and paid to the manager. The scheme studied by Weitzman is linear in the three variables, tentative target, revised target and final output, and it is shown, as in Bonin's paper, that the manager's best answer is to announce the level of output that can be produced with probability r, where r is a simple function of the parameters appearing in the formulae according to which the bonuses are computed.

Other issues are discussed in the above papers, but they all provide a solution to the following problem: assume that the efforts of the manager do not influence the final output but that he is the only agent to possess reliable information about the production possibilities of his firm; this information is represented by a probability distribution $F(\cdot)$. Given a number $0 < r < 1$, the problem is to devise a bonus scheme having the property that the informed agent's optimal response to it is to announce x satisfying $F(x) = r$. In what follows, such a scheme will be called an <u>elicitation scheme</u>.

The purpose of this paper is to characterize all elicitation schemes (Part I), and to study their properties (Part II). Notice that the schemes we are considering are not designed to make the manager work harder, since $F(\cdot)$ is an exogenously given distribution, but rather to give him an incentive to transmit correct information. The intertemporal issue of target adjustments over time is not examined either. Those additional problems are clearly crucial to a complete analysis of managerial incentives but they can be conceptually separated from what will be our topic here. A reader interested in these issues is referred to Weitzman and Snowberger's follow-up [4].

I. Characterization of elicitation schemes

The output y of a firm is a non-negative random variable distributed according to some atomless probability measure $f(\cdot)$, of cumulative distribution $F(\cdot)$. The function $f(\cdot)$ is known to the manager of the firm only.

Assuming that the manager is risk-neutral, the planning authority's (or center's) aim is to set up an incentive scheme that will lead the manager to reveal the level of output y^* satisfying

(1) $F(y^*) = r$, where r is some number between 0 and 1, chosen by the center.

The manager is therefore asked to announce a non-negative scalar x. Production takes place and the actual output y , drawn from $f(\cdot)$ is observed. The manager is then rewarded according to a function H. The characterization of the functions H eliciting $x = y^*$ as the manager's best answer is the object of the following pages.

The functions $f(\cdot)$ will be assumed to be strictly positive, so that the class of admissible mechanisms should be expected to be wider than if this restriction were not imposed. However, no extra mechanisms is actually introduced by this procedure, result that strengthens the characterization theorem. Moreover, this condition allows one to clearly specify the class of mechanisms for which uniqueness of the best strategy is guaranteed. Such uniqueness depends on the uniqueness of the solution to equation (1), ensured by the positivity of $f(\cdot)$.

Step 1 H is a function of y and x , whose integral in y should exist for all $f(\cdot)$ and admit of a maximum in x for all $f(\cdot)$.

Proof y^* cannot be an argument of H since it is never observed by the center.

H has to depend on x if the strategy choice of the manager is to affect his reward at all.

H has to depend on y if the manager's answer is to reveal something about $f(\cdot)$.

It follows that $H = H(x, y)$: H should be a well-defined function of x and y in R^{+2}.

The manager's expected payoff is:

$$p(x) = \int_0^{+\infty} H(x, y)\ f(y)dy\ .$$

This integral should exist for all $f(\cdot)$ and for all x in R^+ and its maximization in x should admit of a solution for all $f(\cdot)$. This establishes step 1.

Consider x_1 and x_2 elements of R^+ and examine the difference:

$$p(x_1) - p(x_2) = \int_0^{+\infty} H(x_1, y)\ f(y)dy - \int_0^{+\infty} H(x_2, y)\ f(y)dy.$$

Truthful elicitation requires that

$$p(x_1) - p(x_2) \geqq 0 \quad \forall x_1, x_2 \in R^+, \quad \forall f(\cdot) \text{ s.t. } F(x_1) = r$$

Define

$$\Delta(x_1, x_2, y) = H(x_1, y) - H(x_2, y)$$

and choose x_1 and x_2 so that $x_1 < x_2$.

<u>Step 2</u> $\forall x_1 \in R^+$, $\forall x_2 \in R^+$, $x_1 < x_2$, $\exists\ G_1(x_1, x_2), G_2(x_1, x_2) \in R$ such that:

$$(2) \begin{cases} (2a) \quad \Delta(x_1, x_2, y) - G_1(x_1, x_2) = 0 \quad \text{a.e. if} \quad y \leqq x_1 \\ (2b) \quad \Delta(x_1, x_2, y) - G_2(x_1, x_2) = 0 \quad \text{a.e. if} \quad y \geqq x_2 \end{cases}$$

Proof: Choose m_i, $M_i \in R$ with $m_i < M_i$ for $i = 1, 2$, and define

$$R_{m_1}(x_1, x_2) = \{y/y \in R^+, y \leqq x_1, \Delta(x_1, x_2, y) \leqq m_1\}$$

$$R_{M_1}(x_1, x_2) = \{y/y \in R^+, y \leqq x_1, \Delta(x_1, x_2, y) \geqq M_1\} .$$

R_{m_2} and R_{M_2} are defined by the same formula with "$y \geqq x_2$" replacing "$y \leqq x_1$".

These four sets are measurable.

Denoting by $\mu(X)$ the measure of the set X, we claim the equivalence of (2) and (3):

(3): $\forall x_1, x_2 \in R, \forall m_i, M_i \in R, m_i < M_i, \forall i \in \{1, 2\}$,

either $\mu(R_{m_i}(x_1, x_2)) = 0$ or $\mu(R_{M_i}(x_1, x_2))$ or both.

It is clear that (2) $\Rightarrow$ (3).

In order to establish the converse, suppose that (3) holds, and define ((m, M) being either pair (m_i, M_i), $i = 1, 2$)

$$\underline{S}(x_1, x_2) = \{m/m \in R, \mu(R_m(x_1, x_2)) > 0\}$$

and

$$\overline{S}(x_1, x_2) = \{M/M \in R, \mu(R_M(x_1, x_2)) > 0\}$$

a) $\forall x_1, x_2 \in R$, $\underline{S}(x_1, x_2) \neq \emptyset$ and $\overline{S}(x_1, x_2) \neq \emptyset$.

Otherwise, H would not be a well-defined function.

b) $\forall x_1, x_2 \in R$, if $m \in \underline{S}(x_1, x_2)$ and $m^1 > m$, then $m^1 \in \underline{S}(x_1, x_2)$.

$\forall x_1, x_2 \in R$, if $M \in \overline{S}(x_1, x_2)$ and $M^1 < M$, then $M^1 \in \overline{S}(x_1, x_2)$.

This follows from the definition of the sets $\underline{S}(x_1, x_2)$ and $\overline{S}(x_1, x_2)$.

It is clear that $\underline{S}(x_1, x_2) \cap \overline{S}(x_1, x_2) \neq \emptyset$.

If, for every x_1, x_2, this interesection has only one point, (2) holds.

If, for a pair x_1°, x_2°, $\exists$ m*, M* with m* < M such that $m^*, M^* \in \underline{S}(x_1^\circ, x_2^\circ) \cap \overline{S}(x_1^\circ, x_2^\circ)$, (3) would be violated since:

$$\mu(R_{m^\circ}(x_1^\circ, x_2^\circ) > 0 \quad \text{and} \quad \mu(R_{M^\circ}(x_1^\circ, x_2^\circ)) > 0$$

with $m^\circ = m^* + 1/3[M^* - m^*]$ and $M^\circ = M^* - 1/3[M^* - m^*]$.

Given the equivalence of (2) and (3), Step 2 is established by showing that (3) holds. The reasoning is by contradiction: suppose that, for a pair x_1°, x_2°, there exist m_i and M_i with $m_i < M_i$, $\mu(R_{m_i}(x_1^\circ, x_2^\circ)) > 0$ <u>and</u> $\mu(R_{M_i}(x_1^\circ, x_2^\circ)) > 0$ for i = 1 or i = 2 or both.

To start with, assume that such inequalities can be found to hold for i = 1 and i = 2 .

Choose a sequence $\{\epsilon_n, n = 1, \ldots\}$ with $\epsilon_n > 0$ $\forall n$, and $\epsilon_n \to 0$ as $n \to +\infty$ and define a sequence $\{f_n, n = 1, \ldots\}$ by

(i) $f_n > 0 \qquad \forall n$

(ii) $F_n(x_1^\circ) = r \qquad \forall n$

(iii) $\int_{R_{m_1}(x_1^\circ, x_2^\circ)} f_n(y)dy = r - \epsilon_n \qquad \int_{R_{m_2}(x_1^\circ, x_2^\circ)} f_n(y)dy = 1 - r - \epsilon_n \qquad \forall n$.

(iv) $1/\epsilon_n f_n(y) = 1/\epsilon_m f_m(y) \quad \forall n, \forall m, \forall y \in R^+ \setminus R_{m_1}(x_1^\circ, x_2^\circ) \cup R_{m_2}(x_1^\circ, x_2^\circ)$.

Truthful elicitation requires that

$$\int_0^{+\infty} \Delta(x_1^\circ, x_2^\circ, y) f_n(y)dy \geq 0 \qquad \forall n .$$

Calling $R^+ \setminus R_{m_1}(x_1^\circ, x_2^\circ) \cup R_{m_2}(x_1^\circ, x_2^\circ) = T_{m_1, m_2}(x_1^\circ, x_2^\circ)$, this can be written:

(4) $$\int_{R_{m_1}(x_1^\circ, x_2^\circ)} \Delta(x_1^\circ, x_2^\circ, y) f_n(y) dy + \int_{R_{m_2}(x_1^\circ, x_2^\circ)} \Delta(x_1^\circ, x_2^\circ, y) f_n(y) dy + \int_{T_{m_1, m_2}(x_1^\circ, x_2^\circ)} \Delta(x_1^\circ, x_2^\circ, y) f(y) dy \geqq 0$$

The three terms of this sum, respectively called A, B and C, are now examined.

$$A \leqq \int_{R_{m_1}(x_1^\circ, x_2^\circ)} m_1 f_n(y) dy = m_1(r - \epsilon_n) .$$

$$B \leqq \int_{R_{m_2}(x_1^\circ, x_2^\circ)} m_2 f_n(y) dy = m_2(1 - r - \epsilon_n).$$

In order to evaluate C, choose a probability function g satisfying:

(i) $g > 0$.

(ii) $g(y) = \frac{1}{2\epsilon_1} f_1(y) \; \left(= \frac{1}{2\epsilon_n} f_n(y)\right), \; \forall \; y \in T_{m_1, m_2}(x_1^\circ, x_2^\circ)$

The integral over any measurable subset of R, of the payoff of a manager whose probability distribution is g, when he announces x_1° or x_2°, is a well defined number

$$\int_{T_{m_1, m_2}(x_1^\circ, x_2^\circ)} H(x_i^\circ, y) \; 1/\epsilon_n f_n(y) dy = 2\int_{T_{m_1, m_2}(x_1^\circ, x_2^\circ)} H(x_i^\circ, y) g(y) dy = Q_i \quad i = 1, 2 .$$

Consequently

$$C \leqq 2\epsilon_n(|Q_1| + |Q_2|) = \epsilon_n Q \text{ with } Q = 2(|Q_1| + |Q_2|) .$$

(4) gives:

$$m_1(r - \epsilon_n) + m_2(1 - r - \epsilon_n) + \epsilon_n Q \geqq 0 \quad \forall n .$$

By letting n go to ∞ ,

(5) $m_1 r + m_2(1 - r) \geqq 0$.

Similarly, by choosing a sequence $\{f_n, n = 1, \ldots\}$ satisfying:

(i) $f_n > 0 \quad \forall n$

(ii) $F_n(x_2^\circ) = r \quad \forall n$.

(iii) $\int_{R_{M_1}(x_1^\circ, x_2^\circ)} f_n(y)dy = r - \epsilon_n$, $\int_{R_{M_2}(x_1^\circ, x_2^\circ)} f_n(y)dy = 1 - r - \epsilon_n$

(iv) $1/\epsilon_n f_n(y) = 1/\epsilon_m f_m(y)$, $\forall n$, $\forall m$, $\forall y \epsilon R^+ \setminus R_{M_1}(x_1^\circ, x_2^\circ) \cup R_{M_2}(x_1^\circ, x_2^\circ)$

One can show that

(6) $M_1 r + M_2(1 - r) \leqq 0$

Subtracting (6) from (5) yields:

$(m_1 - M_1)r + (m_2 - M_2)(1 - r) \geqq 0$, which is inconsistent with $m_1 < M_1$ and $m_2 < M_2$.

If the inequalities stated in (3) had been violated for only one i , the argument would have followed the same lines.

We conclude that (3), and therefore (2), hold. This proves Step 2.

Step 3. H is a separable additive function of x and y .

Proof:

a) Given a sequence $\{x_{2n}, n = 1, \ldots\}$ with $x_{2n} \epsilon R$, $x_{2n} \to +\infty$ as $n \to \infty$, define

$$H_n(x_1, y) = H(x_{2n}, y) - G(x_1, x_{2n}) \quad \forall x_1 \leqq x_{2n}, \forall y \leqq x_1$$

H_n is a separable additive function of x_1 and y in the restricted interval over which it is defined. However, this function does not depend on n as long as $x_{2n} \geqq x_1$. This results from (2a).

Therefore, H satisfying

$$H(x_1, y) = H_n(x_1, y) \quad \forall_n \text{ such that } x_{2n} \geqq x_1, \forall y \leqq x_1$$ is

a well-defined function, and it can be written, like H_n, as a sum of functions of x_1 and y alone.

b) Setting $x_1 = 0$ in (2b) gives

$H(x_2, y) = H(0, y) - G_2(0, x_2)$ if $y \geqq x_2$, which gives the separable form directly. This proves step 3.

To simplify the notations, H is now written:

$$(7)\quad H(x, y) = \begin{cases} A_1(x) + B_1(y) & \text{a.e. if } y \leqq x \\ A_2(x) + B_2(y) & \text{a.e. if } y > x \end{cases}$$

The inequality of the second line is strict so as to ensure consistency at the points $y = x$ without having to write it out explicitly, but this lack of symmetry is immaterial since this is only an "almost everywhere" characterization.

<u>Step 4</u>.

$$(8)\quad [A_1(x_1) - A_1(x_2)]r + [A_2(x_1) - A_2(x_2)](1 - r) = 0 \quad \forall x_1, x_2 \in R$$

<u>Proof</u>: Using (7), $p(x_1) - p(x_2)$ can now be written:

$$p(x_1) - p(x_2) = \int_0^{x_1}[A_1(x_1) + B_1(y) - (A_1(x_2) + B_1(y))]f(y)dy$$

$$+ \int_{x_1}^{x_2}[A_2(x_1) + B_2(y) - A_1(x_2) + B_1(y))]f(y)dy$$

$$+ \int_{x_2}^{+\infty}[A_2(x_1) + B_2(y) - A_2(x_2) + B_2(y))]f(y)dy$$

$$(9)\quad p(x_1) - p(x_2) = [A_1(x_1) - A_1(x_2)]\int_0^{x_1} f(y)dy \; +$$

$$\int_{x_1}^{x_2}[A_2(x_1) + B_2(y) - (A_1(x_2) + B_1(y))]f(y)dy + [A_2(x_1) - A_2(x_2)]\int_{x_2}^{+\infty} f(y)dy$$

Define the sequence $\{f_n, n = 1, \ldots\}$ by

(i) $f_n > 0$ $\quad \forall n$

(ii) $F_n(x_1) = r$ $\quad \forall n$

(iii) $\int_{x_1}^{x_2} f_n(y)dy = \epsilon_n$ $\quad \forall n$

(iv) $1/\epsilon_n f_n(y) = 1/\epsilon_m f_m(y)$ $\quad \forall n, \forall m, \forall y \in [x_1, x_2]$.

Truthful elicitation requires that, if $f = f_n$, $p(x_1) - p(x_2) \geqq 0 \; \forall n$.

The middle term of (9) can be bounded by $\epsilon_n Q$ for some Q, by an argument similar to the one used in evaluating C in step 2.

(9) gives:

$$[A_1(x_1) - A_1(x_2)]r + \epsilon_n Q + [A_2(x_1) - A_2(x_2)](1 - r - \epsilon_n)$$

$$\geqq p(x_1) - p(x_2) \geqq 0 \; .$$

When $n \to \infty$ it follows that

$$(10)\quad [A_1(x_1) - A_1(x_2)]r + [A_2(x_1) - A_2(x_2)](1 - r) \geqq 0$$

Similarly, defining a sequence $g_n > 0$ by

(i) $g_n > 0$ $\quad \forall n$

(ii) $G_n(x_2) = r$ $\quad \forall n$

(iii) $\int_{x_1}^{x_2} g_n(y)dy = \varepsilon_n \qquad \forall n$

(iv) $1/\varepsilon_n g_n(y) = 1/\varepsilon_m g_m(y) \qquad \forall n, \forall m \quad \forall y \in [x_1, x_2]$

would give

(11) $[A_1(x_1) - A_1(x_1)]r + [A_2(x_1) - A_2(x_2)](1 - r) \leqq 0$.

(10) together with (11) imply the equation appearing in the statement of Step 4.

Step 5 a) $B_2(x) - B_1(x) + A_2(x) - A_1(x) = 0$ a.e. in $[0, +\infty[$.

b) $B_2(x) - B_1(x)$ is a.e. a decreasing function of x if B is continuous.

If B is not continuous, statement a) is almost the same but requires a number of definitions that obscure its meaning. Both cases are explicitly dealt with in what follows.

Proof: Call $B(y) = B_2(y) - B_1(y)$, $A(x) = A_2(x) - A_1(x)$.

Using (8), (9) can be written:

(12) $\int_{x_1}^{x_2} B(y)f(y)dy + A(x_2)\int_{x_1}^{x_2} f(y)dy \geqq 0$ if $F(x_1) = r$

(13) $\int_{x_1}^{x_2} B(y)f(y)dy + A(x_1)\int_{x_1}^{x_2} f(y)dy \leqq 0$ if $F(x_2) = r$

Given a sequence $\{\varepsilon_n, n = 1, \ldots\}$, $\varepsilon_n > 0 \ \forall n$, $\varepsilon_n \to 0$ as $n \to \infty$, define, for $m \in R$, $R_m(x_1, x_2) = \{y/y \in [x_1, x_2]$ and $B(y) \leqq m\}$ and $\underline{S}(x_1, x_2) = \{m/m \in R$ and $\mu(R_m(x_1, x_2)) > 0\}$.

By an argument similar to the one presented in step 2,

a) $S(x_1, x_2) \neq \emptyset$

b) If $m \in \underline{S}(x_1, x_2)$ and $m^1 > m$, then $m^1 \in \underline{S}(x_1, x_2)$.

Call $m^*(x_1, x_2) = \inf m, m \in \underline{S}(x_1, x_2)$. $m^*(x_1, x_2)$ exists, otherwise H would not be well defined on $[x_1, x_2]$.

Define the sequence $\{f_n, n = 1, \ldots\}$ by:

(i) $f_n > 0 \quad \forall n$

(ii) $F_n(x_1) = r, \quad F_n(x_2) = r + s$ with $s > 0 \quad \forall n$

(iii) $f_n(y) = \epsilon_n \quad \forall\, y \in [x_1, x_2] \setminus R_{m^*(x_1, x_2) + \epsilon_n} \quad \forall n$

(12) gives

$$p(x_1) - p(x_2) \leqq \int\limits_{R_{m^*(x_1, x_2) + \epsilon_n}} [m^*(x_1, x_2) + \epsilon_n] f(y)dy$$

$$+ \int\limits_{[x_1, x_2] \setminus R_{m^*(x_1, x_2) + \epsilon_n}} B(y) f(y) dy \qquad + \; A(x_2) \int_{x_1}^{x_2} f(y) dy$$

$$\text{(ii)} \Rightarrow \int\limits_{R_{m^*(x_1, x_2) + \epsilon_n}} f(y) dy \leqq \int_{x_1}^{x_2} f(y) dy = s$$

$$\text{(iii)} \Rightarrow \int\limits_{[x_1, x_2] \setminus R_{m^*(x_1, x_2) + \epsilon_n}} B(y) f(y) dy \leqq \epsilon_n \int_{x_1}^{x_2} |B(y)| dy = \epsilon_n Q^1 .$$

since $\int_{x_1}^{x_2} |B(y)| dy$ is a well-defined number, designated by Q^1.

Therefore:

$$p(x_1) - p(x_2) \leqq (m^*(x_1, x_2) + \epsilon_n) s + \epsilon_n Q^1 + A(x_2) s .$$

After division by s, and an obvious change of notation, it follows that

(14) $0 \leqq m^*(x_1, x_2) + A(x_2) + \epsilon'_n \quad \forall n$

with $\epsilon'_n \to 0$ as $n \to +\infty$.

Similarly, define, for $M \in R$, $R_M(x_1, x_2) = \{y/y \in [x, x_2]$ and $B(y) \geqq M\}$ and $\overline{S}(x_1, x_2) = \{M/M \in R$ and $\mu(R_M(x_1, x_2)) > 0\}$. As before,

a) $\overline{S}(x_1, x_2) \neq \emptyset$

b) If $M \in \overline{S}(x_1, x_2)$ and $M^1 < M$, then $M^1 \in \overline{S}(x_1, x_2)$.

Call $M^*(x_1, x_2) = \sup M, M \in \overline{S}(x_1, x_2)$. $M^*(x_1, x_2)$ exists, since H is well-defined on $[x_1, x_2]$.

It could then be derived that:

$$(15) \quad 0 \geqq M^*(x_1, x_2) + A(x_1) + \epsilon''_n \quad \forall n$$

$$\text{with } \epsilon''_n \to 0 \text{ as } n \to +\infty .$$

Perform the change of notation $x_2 = x$ in (14) and $x_1 = x$ in (15).

By letting n go to infinity,

(14) gives: $m^*(x_1, x) + A(x) \geqq 0 \quad \forall x_1 \leqq x$

(15) gives: $M^*(x, x_2) + A(x) \leqq 0 \quad \forall x \leqq x_2$

Together we get:

$$(16) \quad -M^*(x, x_2) \geqq A(x) \geqq -m^*(x_1, x). \quad \forall x_1 \leqq x \leqq x_2$$

This is possible only if $m^*(x_1, x) \geqq M^*(x, x_2) \quad \forall x_1 \leqq x \leqq x_2$ and this in turn implies that B is non-increasing almost everywhere.

If x is a point of continuity of B, when $x_1 \to x$, $m^*(x_1, x) \to B(x)$ and when $x_2 \to x$, $M^*(x, x_2) \to B(x)$.

(16) then yields: $A(x) + B(x) = 0$.

If x is not a point of continuity of B, it is sufficient that A(x) satisfies (16). The sufficiency of (16) is seen by going back to (12):

$$\int_{x_1}^{x_2} B(y)f(y)dy + A(x_2)\int_{x_1}^{x_2} f(y)dy \geqq m^*(x_1, x_2)\int_{x_1}^{x_2} f(y)dy + A(x_2)\int_{x_1}^{x_2} f(y)dy$$

$$= [m^*(x_1, x_2) + A(x_2)]\int_{x_1}^{x_2} f(y)dy ,$$ which is non-negative

if (16) holds.

(13) could be checked in a similar way.

Finally, notice that uniqueness of the best strategy is guaranteed by the choice of a strictly monotone B. If B were constant a.e. on a segment $[c, d]$, A would also have that property, and the payoff would be insensitive to changes of strategies in $[c, d]$. Conversely, if strict monotonicity holds, the above inequalities are strict so that deviation from the truth leads to a strict loss in expected payoff.

All these conclusions are gathered in the following proposition:

Proposition: A necessary and sufficient condition for a scheme H to elicit x^*, solution of "$F(x) = r$", as best answer is that H satisfies:

$$H(x, y) = \begin{cases} A_1(x) + B_1(y) & \text{a.e. if } y \leqq x \\ A_2(x) + B_2(y) & \text{a.e. if } y > x \end{cases}$$

with $(A_1(x_1) - A_1(x_2))r + (A_2(x_1) - A_2(x_2))(1 - r) = 0 \quad \forall x_1 \; \forall x_2$.

and $-M^*(x, x_2) \geqq A(x) \geqq -m^*(x_1, x) \quad \forall\, x_1 \leqq x \leqq x_2$ where

$m^*(x_1, x)$ and $M^*(x, x_2)$ are defined in step 5.

If $B_2(x) - B_1(x)$ is taken to be continuous, this last condition can be written:

$B_2(x) - B_1(x)$ does not increase.

(17) $B_2(x) - B_1(x) + A_2(x) - A_1(x) = 0 \quad \forall x$.

Uniqueness of the best strategy is guaranteed by enforcing strict monotonicity of $B_2(x) - B_1(x)$.

Remark

This paper differs from Weitzman's in that output is here restricted to be non-negative while it may be any real number in [1]. In addition, Weitzman's scheme elicits x^* satisfying $F(x^*) = 1 - r$, while we are concerned with the solution x^* of "$F(x) = r$". Both differences are inessential.

II. Properties of the elicitation schemes

1) If the distribution function $f(\cdot)$ has mass-points, F is discontinuous and there might be no solution to "$F(x) = r$". On the other hand, since the schemes characterized in the proposition may exhibit discontinuities, the Stieltjes integral giving the manager's expected payoff may not be well defined. Given any point of discontinuity of H , one can find a function F discontinuous at the same point, for which the expected payoff cannot be computed. It follows that for every x , $H(x, y)$ should be a continuous function of y . This is possible for each of the intervals $[o, x[$ and $]x, +\infty[$ and is guaranteed at $y = x$ by condition (17). The "almost everywhere" statements of the proposition should be replaced by "everywhere" statements. The class of elicitation schemes is then somewhat narrower. If

a discontinuity of F appears at a point x^* such that $F(x^*) > r$, and $F(x) < r \quad \forall x < x^*$, x^* is nevertheless the manager's best answer, so that the nonexistence of a solution to "$F(x) = r$" does not prevent useful information to be transmitted to the center. Sensitivity analysis only (operating the scheme for different values of r) would reveal to the center the existence of the discontinuity. This is important to notice because it is not true of all elicitation schemes that a manager confronted with a "multiple" questionaire, will still have an incentive to give the correct answer to each of the individual questions.

2) <u>If the distribution function $f(\cdot)$ is zero over some range</u>, F has a flat section, and there exists a continuum of solutions to "$F(x) = r$". It is likely though that the center would be particularly interested in the highest point of this range, corresponding to the greatest output. However, no scheme of the class given by the proposition could induce the manager to choose that point. A small variation of r could however reveal to the center the existence and the length of the flat section. As before, sensitivity analysis seems necessary to the obtention of the extra information.

3) Given a non-negative number λ, if H_1 and H_2 are elicitation schemes, so are $H_1 + H_2$ and λH_1.

4) <u>Existence of bounded and individually-rational specifications</u>.

In the scheme presented by Weitzman, the transfers between center and manager may be quite large if the difference between y and x happens to be important. It is of interest to investigate the existence of elicitation schemes whose maximum cost is bounded above and below, because they guarantee that neither center nor manager will go bankrupt, if they have limited resources to start with.

In addition to satisfying the conditions of the proposition, an additional requirement is therefore imposed on the elicitation scheme, namely:

$$N \leqq H(x^*, y) \leqq M . \quad \forall y .$$

However, since the designer does not know beforehand what x^* will be, it should be case that

$$N \leqq H(x, y) \leqq M \quad \forall x, \forall y .$$

which implies:

$$(18)\ N \leqq A_1(x) + B_1(y) \leqq M \quad \forall x, \forall y \leqq x.$$

setting $y = 0$, this yields

$$N - B_1(0) \leqq A_1(x) \leqq M - B_1(0) \quad \forall x .$$

Therefore, <u>$A_1(\cdot)$ is bounded below and above</u>; (a)

Next, one computes, with the help of (8), in which $x_1 = x$ and $x_2 = 0$,

$$A(x) = A_2(x) - A_1(x) = \frac{-1}{1-r} A_1(x) + \frac{r}{1-r} A_1(0) + A_2(0)$$

Since $A(\cdot)$ is non-decreasing, <u>$A_1(\cdot)$ should be non-increasing</u>; (b) Given the choice of a function $A_1(\cdot)$ satisfying (a) and (b), one chooses B_1 satisfying (18).

$$(18) \Rightarrow N - A_1(x) \leqq B_1(y) \leqq M - A_1(x) \quad \forall y, \forall x \geqq y .$$

Noting that $-A_1(\cdot)$ is an increasing function and calling m_o the highest lower bound of $A_1(\cdot)$, we get:

$$(19) \quad N - M_o \leqq B_1(y) \leqq M - A_1(y) \qquad \forall y .$$

In order for this to be possible, A_1 should then satisfy:

$$N - M_o \leqq M - A_1(y) \qquad \forall y$$

$$\text{or} \quad A_1(y) \leqq m_o + M - N$$

i.e., the range of variation of $A_1(.)$ should not be greater than $M - N$.

Then $B_1(y)$ can be chosen so as to satisfy (19) .

$B_2(y)$ is finally given by (17). This completes the construction and establishes the existence of bounded schemes.

A particular case is when no upper bound is put on the payment ($M = \infty$) but individual rationality is required for the manager ($N = 0$). The preceding construction shows that individually rational specifications exist.

Since M and N are arbitrary (except for the requirement $M > N$), the cost to the center of running an elicitation scheme can be made as small as one wishes. One should be careful though since a choice of a small difference $M - N$ implies a relative unresponsiveness of the manager's reward to his efforts to find out what x^* is. The loss he incurs from making a "mistake" becomes less and less serious and may destroy his incentive to evaluate F correctly as soon as minimal costs of information gathering are introduced, as they should in a more realistic analysis of his behavior.

5) Schemes involving the transfer of less information.

In the derivation of Part I, it was assumed that the center would eventually know y, as well as keep the manager's answer on record, so that the bonus could be expressed as a function of both x and y.

However, schemes based on less information are conceivable: assume for instance that the center only observes whether y is above x, without ever knowing the precise value of y. Conversely, the center could observe y and "remember" only whether the announced x was above or below y. The bonus scheme would involve functions of x only in the first case, and functions of y only in the second case. The functions could differ for $y \leq x$ and $y > x$. The examination of the conditions stated in the propositions reveals however that such schemes could not elicit x^*.

The dependence of the bonus on x and y, stated in Step 1, should therefore be non-trivial where "trivial" means "only through the cut-off condition $y = x$".

6) Case of restricted families of distributions.

If $f(\cdot)$ is known to belong to a certain class of probability distributions, a richer class of elicitation schemes can be found. In order to illustrate this comment, suppose that the designer knows that $f(\cdot)$ is a uniform distribution over $[0, t]$, where t is some (unknown to him) parameter.

The manager's payoff can be written:

$$p(x) = \int_0^t H(x, y)\ f(y)dy = \frac{1}{t}\int_0^t H(x, y)dy.$$

x^* is given by $x^* = rt$, and the loss from announcing $x \neq x^*$

is the negative of

$$p(x^*) - p(x) = \frac{1}{t}\int_0^t [H(x^*, y) - H(x, y)]dy$$

Denoting $\int H(x, y)dy = h(x, y)$, elicitation requires that:

$$[h(x^*, y) - h(x, y)]_0^t \geqq 0 \qquad \forall x\ \forall t$$

$\Leftrightarrow$ (20) $h(rt, t) - h(rt, 0) - h(x, t) + h(x, 0) \geqq 0 \quad \forall x\ \forall t$.

(20) is the only necessary and sufficient condition that H should satisfy in order to lead to the elicitation of x^* .

For example, if $h(x, 0) = 0$, (20) reduces to

$h(rt, t) - h(x, t) \geqq 0$, which means that

as a function of x , $h(x, t)$ should be maximized at $x = rt$, for all t .

Otherwise, given any arbitrary function $h(\cdot\ , 0)$, it is enough to choose $h(x, t)$ satisfying (20) for every t .

7) Case of differentiable elicitation schemes.

If the search for elicitation schemes is restricted to payment functions that are differentiable on each of the subsets of R^{2+} defined by $\{(x, y) \in R^{2+} / y < x\}$ and $\{(x, y) \in R^{2+} / y > x\}$, the characterization proof takes a somewhat different form:

For simplicity, it is assumed that H satisfies conditions allowing to take the derivative under the integral sign appearing in the definition of $p(x)$ in order to evaluate the derivative of $p(x)$.

Writing $\begin{cases} H(x, y) = H_1(x, y) & \text{if } y < x \\ H(x, y) = H_2(x, y) & \text{if } y \geqq x \end{cases}$, $p(x)$ is given by:

$$p(x) = \int_0^x H_1(x, y) f(y) dy + \int_x^{+\infty} H_2(x, y) f(y) dy .$$

Then

$$p'(x) = \int_0^x \frac{\partial H_1(x, y)}{\partial x} + \int_x^{+\infty} \frac{\partial H_2(x, y)}{\partial x} + H_1(x, x) f(x) - H_2(x, x) f(x).$$

Elicitation requires that

$$p'(x^*) = 0 \quad \text{where } x^* \text{ satisfies } F(x^*) = r .$$

It would be to lengthy to develop the derivation of the conditions (i) - (ii) listed below. Such derivation would closely follow the steps indicated in the non-differentiable case.

(i) $H_1(x, x) - H_2(x, x) = 0 \qquad \forall x .$

(ii) $\dfrac{\partial H_1(x, y)}{\partial x} = a_1(x) , \quad \dfrac{\partial H_2(x, y)}{\partial x} = a_2(x) \qquad \forall x, \forall y .$

(iii) $a_1(x) r + a_2(x)(1 - r) = 0 \qquad \forall x .$

Integration of (ii) in x would then yield:

$$\text{(a)} \quad H(x, y) = \begin{cases} A_1(x) + B_1(y) & \text{if } y \leqq x \\ A_2(x) + B_2(y) & \text{if } y > x \end{cases}$$

with (b) $\quad A_1(x) - A_2(x) + B_1(x) - B_2(x) = 0 \qquad \forall x$

and (c) $\quad A'_1(x) r + A'_2(x)(1 - r) = 0 \qquad \forall x.$

These conditions are necessary. They allow to rewrite $p'(x)$ as:

$$p'(x) = A'_1(x)F(x) + A'_2(x)(1 - F(x)) \quad \text{, and then}$$

$$= A'_1(x) \left[F(x) - \frac{r}{1-r} (1 - F(x))\right]$$

$$= \frac{A'_1(x)}{1 - r} \left[F(x) - r\right] .$$

$F(x) - r$ is non-positive for $x \leqq r$, and non-negative for $x \geqq r$. In order for a maximum to be attained at x^*, it should be the case that $p'(x) \geqq 0$ for $x \leqq x^*$ and $p'(x) \leq 0$ for $x \geq x^*$. This is achieved if and only if $A'_1(x^*) \leqq 0$. Since x^* can be any number, a necessary and sufficient condition for the extremum guaranteed by (a), (b) and (c) to be a maximum is that $A'_1(x) \leqq 0$, $\forall x \in R$. This completes the characterization in the differentiable case.

Remark

For the direct elicitation of probabilities, converse problem of the one we study here, the reader is referred to Savage [1].

References

[1] Bonin, J. P., "On the Design of Managerial Incentive Structures in a Decentralized Planning Environment". American Economic Review, 66 (1976).

[2] Fan, L.S., "On the Reward System". American Economic Review, 65 (1975).

[3] Savage, L. J., "Elicitation of Personal Probabilities and Expectations". Journal of the American Statistical Association, December 1971.

[4] Snowberger, W. "The New Soviet Incentive Model: A Comment". Bell Journal of Economics, Autumn 1977.

[5] Weitzman, Martin L., "The New Soviet Incentive Model". Bell Journal of Economics, 7 (Spring 1976).