IMPERFECT PERSONAL INFORMATION AND THE DEMAND REVEALING PROCESS: A SAMPLING APPROACH.

JERRY GREEN AND JEAN-JACQUES LAFFONT*

I. INTRODUCTION

Public decision processes involving large numbers of agents face a serious difficulty whenever there are costs involved in the precise reporting of individuals' preferences. This obstacle to efficient decision-making exists independently of whether there is any incentive problem regarding the agents' desires to report their true preferences. It may arise either because voting, or preference revelation more generally, is a costly act, or because the individual does not know his own true tastes precisely but can learn them more accurately through costly introspection or information acquisition.

Recently, progress has been made on the incentive question first set forth by K. Wicksell (1896) and further expounded by R. Musgrave (1959) and P. Samuelson (1954), among others.[1] The goal of this article is to study whether the solution to this problem proposed by E.H. Clarke (1971) and T. Groves (1973)[2]

*Harvard University, Cambridge, U.S.A. and Laboratorie d'Econometrie de l'Ecole Polytechnique, C.N.R.S., Paris, France respectively. A preliminary version of this paper was presented to the 11th NSF-NBER Conference on Bayesian Econometrics, Rochester, 1975.

[1]Early positive results were attained by H. Bowen [1943], E.A. Thompson [1967], and M. Kurz [1974]. But see P. Bohm [1972] for some other evidence on the severity of the problem in practice.

[2]See J. Green and J.J. Laffont [1976] for detailed comments on the historical background.

can be modified through the use of sampling techniques to overcome the weakness of the individuals' incentives to respond accurately in large numbers situations. We give a qualified affirmative answer to this question. For a particular set of stochastic specifications it can be shown that by using the optimal sample size an expected social outcome superior to that attainable by sampling the entire population can be achieved. However the solution is still strictly worse than perfect and costless individual information would allow.

We consider a single public project, fixed in size. The preferences of each economic agent depend on the acceptance or rejection of this project and on the level of monetary transfers he receives, if any. Following Clarke and Groves, we will assume these take the additively separable form:

$$u_i = v_i + t_i \quad \text{if the project is accepted}$$

$$= t_i \quad \text{if the project is rejected}$$

where t_i is the monetary transfer. Since the cost of the project is imputed to the individuals, the willingness to pay, v_i, may be negative. Even though the project is desirable in itself, individuals may not be willing to bear their assigned shares of the costs.

Let w_i be the announced willingness to pay for agent i. The notation

$$w_{-i}=(w_1, \ldots, w_{i-1}, w_{i+1}, \ldots, w_N)$$

will be employed as a convenient shorthand.

Clarke (1971) has given a social-decision mechanism that is particularly attractive since it causes each individual to announce his true v_i, independently of his beliefs concerning the willingness to pay others. The project is accepted whenever $\sum_i w_i \geqslant 0$, and individual i pays a tax equal to $\left| \sum_{j \neq i} w_j \right|$, if his answer changes the sign of the sum, that is,

$$\text{if} \quad \sum_i w_i \geq 0 \quad \text{and} \quad \sum_{j \neq i} w_j < 0$$

$$\text{or if} \quad \sum_i w_i < 0 \quad \text{and} \quad \sum_{j \neq i} w_j \geq 0.$$

Therefore, agent i's gain in utility over the status quo situation is:

$$\begin{array}{llll}
v_i & \text{if} & \sum_i w_i \geq 0 & \text{and} \quad \sum_{j \neq i} w_j \geq 0 \\
v_i + \sum_{j \neq i} w_j & \text{if} & \sum_i w_i \geq 0 & \text{and} \quad \sum_{j \neq i} w_j < 0 \\
- \sum_{j \neq i} w_j & \text{if} & \sum_i w_i < 0 & \text{and} \quad \sum_{j \neq i} w_j \geq 0 \\
0 & \text{if} & \sum_i w_i < 0 & \text{and} \quad \sum_{j \neq i} w_j < 0
\end{array}$$

It is easy to see that for agent i (i=1, . . . ,N), any choice w_i other than the truth v_i can only lead to a smaller payoff without having any potential for gain.

One weakness of all public decision making processes remains a problem in this mechanism as well. When the number of individuals increases, each has a diminishing incentive to participate in the preference elicitation process. Moreover, if his own tastes are unknown to him and can be discovered only at a cost in real resources, he may not find it in his best interest to do so. Given that he participates, his response will not in general be a valid representation of his tastes.[3]

To circumvent this difficulty we introduce a sampling version of the Clarke mechanism in which we ask only a sample of the population in order to keep the strength of the incentive high enough. But a sample implies a sampling error which can lead to wrong decisions.

In Section II, we explain why the strength of the incentive decreases with the size of the sample. The individual tradeoff between the information cost associated with an individual's search for his true evaluation and the risk associated with a false answer is formalized in Section III. Assuming that the decision-maker is a Bayesian agent and that the population consists of rational expected utility maximizing individuals, we obtain in Section IV the optimal sample size for certain stochastic specifications. Section V shows that the informational value of the procedure is positive. Concluding remarks are gathered in Section VI.

II. STRENGTH OF INCENTIVE

We consider an agent who is in a sample of size n + 1 and who is asked, through a Clarke mechanism, to reveal his evaluation of a given public project, the cost of which is assumed to be zero for simplicity of notation. Let x denote the sum of the answers of the n other agents. From the definition in Section I we see that agent i's gain, $B(v_i, w_i)$, of saying the truth v_i instead of w_i is:

$$
\begin{array}{ll}
0 & \text{if } v_i + x \geq 0 \text{ and } w_i + x \geq 0 \\
 & \text{or if } v_i + x < 0 \text{ and } w_i + x < 0; \\
v_i + x & \text{if } v_i + x \geq 0 \text{ and } w_i + x < 0; \\
-v_i - x & \text{if } v_i + x < 0 \text{ and } w_i + x \geq 0.
\end{array}
$$

The value of 0 occurs when v_i and w_i lead to the same social decision. The value of v_i+x occurs if agent i's statement of w_i defeats the project when the truth, v_i, would not have done so. The value of $-v_i-x$ arises if agent i's statement of w_i leads to acceptance of the project when the truth, v_i, would have killed it.

[3]Groves [1973] has shown that there is a whole family of mechanisms for which the truth is a dominant strategy. They differ from the Clarke mechanism by arbitrary functions of the others' answers. In Green and Laffont [1975 a] we have proved that this family characterizes the mechanisms which are successful and for which the truth is a dominant strategy.

Let G_n^i (x) be agent i's subjective distribution function over the sum of the answers of the n other agents. If agent i is risk neutral, his expected gain of saying v_i instead of w_i is, for $v_i \geqslant 0$:

$$E_n\, B(v_i, w_i) \text{ is } \int_{-v_i \leqslant x < -w_i} (v_i + x)\, d\, G_n^i(x) \qquad \text{if} \qquad v_i > w_i$$

$$\text{or} \qquad - \int_{-w_i \leqslant x < -v_i} (v_i + x)\, d\, G_n^i(x) \qquad \text{if} \qquad v_i < w_i .$$

Observe that

$$E_n\, B(v_i, w_i) \quad < \quad [v_i - w_i] \;.\; \Pr_n [-v_i \leqslant x < -$$

$$\text{if } v_i > w$$

$$< \quad [w_i - v_i] \;.\; \Pr_n [-w_i \leqslant x < -v_i]$$

$$\text{if } v_i < w$$

An analogous expression can be derived for $v_i < 0$. Clearly, as n grows, the probability that x belongs to a fixed interval $[-w_i, -v_i]$ (or $[-v_i, -w_i]$) decreases for n large enough. For example, if the subjective probability distribution of x is normal with mean zero and variance n, we have:

$$\Pr_n \left[-w_i \le x < -v_i\right] = \frac{1}{\sqrt{2\pi}} \int_{-w_i}^{-v_i} \frac{e^{-\frac{x^2}{2n}}}{\sqrt{n}}\, dx$$

$$\psi_n(x) = \frac{e^{-\frac{x^2}{2n}}}{\sqrt{n}}$$

converges uniformly to zero as n goes to infinity. Consequently, $\Pr_n[-w_i \leqslant x < -v_i]$ converges to zero as n tends to infinity.

Therefore, the strength of incentive, measured as the expected gain of saying the truth instead of any other fixed answer w_i, goes to zero as n goes to infinity.

III. *INDIVIDUAL BEHAVIOR*

We suppose now that the agent does not know his own evaluation exactly, but that he can acquire costly information about the project itself and about his own preferences. This uncertainty is here formalized in the following way: each agent has a prior belief concerning his own evaluation given by a normal distribution $N(0,r^2)$. His information gathering[4] process is simplified to the extreme; if he spends an amount c he discovers his true evaluation v. Otherwise he gives an answer which minimizes his expected regret, that is, the expected shortfall in utility attained versus that attained under prefect information. We assume that the ability to process information differs among the agents and that there exists in the sample a distribution function F(c)[5] which describes the distribution of necessary costs to discover the truth.

Let v_i be the true evaluation of agent i: if he buys information at the cost c_i, he answers v_i, since we know from the Clarke mechanism (see Clarke [1971] Green and Laffont [1975a]), that the response v_i dominates any other response w_i.

[4]In Green and Laffont [1975b], we chose a more complex information gathering process in which an agent has the opportunity to draw an observation w from a probability distribution whose mean is the true value and whose variance decreases with the expense incurred. Then, the agent minimizes his expected regret using his posterior distribution. This model does not permit closed form solutions and requires simulations which prevent a clear exposition of the main ideas.

[5]We approximate this discrete distribution by a continuous distribution independently of the sample size.

Consider now an agent i who does not buy information. Let us recall that x be the sum of the answers of all other agents

$$x = \sum_{j \neq i} w_j$$

The regret of saying w_i when the truth is v_i can be expressed as:

$$R_n(v_i, w_i) = \int_{w_i \leqslant -x < v_i} |v_i + x| \; g_n^i(x)\, dx$$

if the agent is underreporting his benefits where $g_n^i(x)$ is agent i's subjective probability density over the sum of the others' answers. An analogous expression where the limits of integration are $-w_i \leqslant x < v_i$ applies if the agent is overreporting. We assume that agent's beliefs are restricted to the family of normal distributions. The mean and the variance depend on his expectations about who buys information. Suppose that he believes that an agent who does not buy information [their number is, denoted by n_1] gives a non-random answer μ. Let $k = n_1 \mu$. Then $x = k + \sum_{i=1}^{n_2} v_i$, where n_2 is the number of other agents in the sample who buy information. Agent i believes that the others' true evaluations are independent and identical to his expectations concerning his own evaluation. Consequently,

$$E\,x = k$$

$$\operatorname{Var} x = n_2 r^2$$

Therefore, $$g_n^i(x) = \frac{1}{r\sqrt{2\pi n_2}}\, e^{-\frac{1}{2}\frac{(x-k)^2}{n_2 r^2}}$$

Proposition 1: Whatever his expectations about the answers of the agents who do not buy information, if an agent i does not buy information, he answers the mean of his prior, 0. Consequently, he is led to believe that the other agents who do not buy information also answer 0, i.e, k = 0.

Proof: Consider agent i who does not buy information. His optimal response is determined by the minimization of expected regret given v_i, i.e,:

$$\min_{w_i} \int_{-\infty}^{+\infty} R_n(v_i, w_i) \; f(v_i) \; dv_i$$

where $f(v_i)$ is his prior density function, i.e., $\frac{1}{r\sqrt{2\pi}} \; e^{-\frac{v_i^2}{2r^2}}$

Let us first compute $R_n(v_i, w_i)$

If $v_i > w_i$, then:

$$R_n(v_i, w_i) = \int_{-v_i}^{-w_i} (v_i + x) \frac{1}{r\sqrt{2\pi n_2}} e^{-\frac{1}{2} \frac{(x-k)^2}{n_2 r^2}} dx$$

$$= (v_i + k) \int_{-v_i}^{-w_i} \frac{1}{r\sqrt{2\pi n_2}} e^{-\frac{1}{2} \frac{(x-k)^2}{n_2 r^2}} dx$$

$$+ \int_{-v_i}^{-w_i} \frac{(x-k)}{r\sqrt{2\pi n_2}} e^{-\frac{1}{2} \frac{(x-k)^2}{n_2 r^2}} dx =$$

$$(v_i + k) \left[\rho\left(\frac{-w_i - k}{r\sqrt{n_2}}\right) - \rho\left(\frac{-v_i - k}{r\sqrt{n_2}}\right) \right] +$$

$$\frac{r\sqrt{n_2}}{\sqrt{2\pi}} \left[e^{\frac{-(v_i+k)^2}{2n_2 r^2}} - e^{\frac{-(w_i+k)^2}{2n_2 r^2}} \right]$$

$$\text{where} \quad \rho(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{u^2}{2}} du$$

The same result obtains for $v_i < w_i$. Therefore, the expected regret is:

$$(1) \int_{-\infty}^{+\infty} \left\{ (v_i+k)\left[\rho\left(\frac{-w_i-k}{r\sqrt{n_2}}\right) - \rho\left(\frac{-v_i-k}{r\sqrt{n_2}}\right)\right] + \frac{r\sqrt{n_2}}{\sqrt{2\pi}}\left[e^{\frac{-(v_i+k)^2}{2n_2r^2}} - e^{\frac{-(w_i+k)^2}{2n_2r^2}}\right]\right\} \frac{e^{-\frac{v_i^2}{2r^2}}}{r\sqrt{2\pi}} \, dv_i$$

We differentiate with respect to w_i and obtain, the first-order condition,

$$\frac{e^{-\frac{(w_i+k)^2}{2n_2r^2}}}{\sqrt{2\pi n_2}} \int_{-\infty}^{+\infty} (w_i + k - v_i - k) \frac{e^{-\frac{v_i^2}{2r^2}}}{r\sqrt{2\pi}} \, dv_i = 0 .$$

Therefore, $w_i = 0$

Q.E.D.

Therefore, we know that if he buys information, the agent i answers the truth v_i and he answers 0 if he does not buy information. The remaining problem is: how does he decide whether or not to buy information?

He will buy information if the expected regret when answering 0 is larger than his cost, c_i. We take an equilibrium theoretic point of view in this matter, and assume that each individual makes a decision, based on his own cost, in the belief that all other individuals are performing the same calculation relative to their own costs. An equilibrium situation will be one in which these beliefs are verified by the actual decisions made. Let F(.) denote the distribution of these costs, c.

If we let k = 0 in (1), we obtain:

$$E R_n(v_i, 0) = 2\int_0^{\infty} \left[v_i \left[\frac{1}{2} - \rho\left(-\frac{v_i}{r\sqrt{n_2}}\right)\right] + \frac{r\sqrt{n_2}}{\sqrt{2\pi}} \left(e^{-\frac{v_i^2}{2n_2r^2}} - 1\right)\right] \frac{e^{-\frac{v_i^2}{2r^2}}}{r\sqrt{2\pi}} \, dv_i$$

$$= \frac{r}{\sqrt{2\pi}} \left[\sqrt{n_2+1} - \sqrt{n_2}\right]$$

as may be seen by complex but straight forward manipulations.

Since all agents face the same problem, there must be a cut-off point c^* after which it is not worthwhile to buy information. Then,

$$n_2 = n \int_0^{c^*} dF(c) = nF(c^*)$$ [6]

Therefore, c^* is a solution of the equation

$$(2) \qquad \frac{r}{\sqrt{2\pi}} \left[\sqrt{nF(c^*)+1} - \sqrt{nF(c^*)} \right] = c^*$$

The agent must obtain the cut-off point c^* and then compare his own c_i to c^*. If $c_i > c^*$, he does not buy information and answers 0. If $c_i < c^*$, he buys information, he obtains the truth v_i, and he answers by revealing it.

[6] In fact n_2 is a random variable with a binomial distribution $[n, F(c)]$ so that

$$E R_n (v_i, 0) = E[E R_n (v_i, 0 / n_i)]$$

$$= \frac{r}{\sqrt{2\pi}} \sum_{n_2=0}^{n} \left[\sqrt{n_2+1} - \sqrt{n_2} \right]$$

$$\frac{n!}{n_2!\,(n-n_2)!} F(c)^{n_2} [1-F(c)]^{n-n_2}$$

We approximate $E[\sqrt{n_2+1} - \sqrt{n_2}]$ by

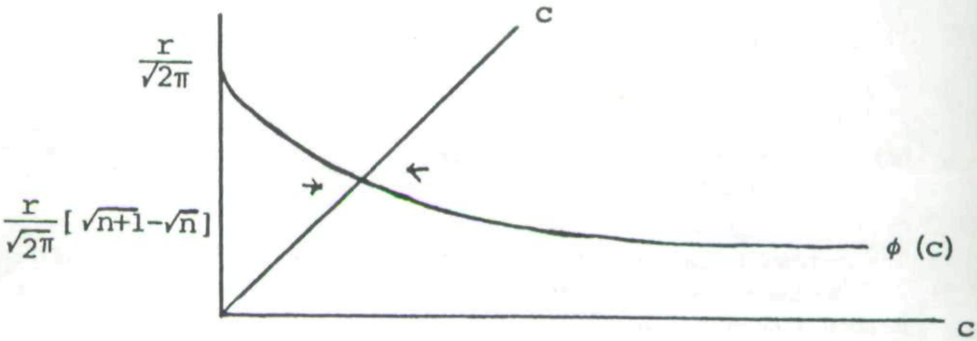
$$\sqrt{E n_2 + 1} - \sqrt{E n_2}$$

Proposition 2: There is a unique equilibrium cut-off point.

Proof: Let $\phi(c) = \frac{r}{\sqrt{2\pi}}\left[\sqrt{nF(c)+1} - \sqrt{nF(c)}\right]$

$$\frac{d\phi}{dc}(c) = \frac{r}{2\sqrt{2\pi}}\left[\frac{n\,f(c)}{\sqrt{n\,F(c)+1}} - \frac{n \cdot f(c)}{\sqrt{n\,F(c)}}\right] < 0$$

To determine the equilibrium cut-off point we have the following diagram



Therefore, there is a unique cut-off point. Q.E.D.

Moreover, the cut-off point is stable in the following sense. E $R_n(v_i, 0)$ is a decreasing function of c. To obtain the equilibrium c^*, the agent can start from c, compute a number n_2 of agents buying information by $n_2 = n\int_0^c dF(c)$. Then, if $ER_n(v_i,0) > c$, the expected cost of not buying information is larger than c, suggesting that the cut-off point is larger, etc. . .

Proposition 3: The cut-off point goes to zero as the sample size goes to infinity; when the uncertainty of the agent increases (r^2 increases) the cut-off point increases, i.e., a larger proportion of agents is buying information.

Proof: It is easy to check by differentiation of (2) that

$$\frac{dc^*}{dr} > 0$$

Suppose that c^* converges to $\bar{c} \neq 0$ when n goes to infinity; then the left hand side of (2) goes to zero, a contradiction since the right hand side goes to $\bar{c}$.

Q.E.D.

In order to obtain a closed form relation between the equilibrium cut-off point and the sample size, we first use the approximation:

$$\sqrt{n_2 + 1} - \sqrt{n_2} \approx \frac{1}{2\sqrt{n_2+1}} \qquad \text{for } n_2 \text{ large}.$$

Equation (2) then becomes:

$$\frac{r}{2\sqrt{2\pi}} \quad \frac{1}{\sqrt{nF(c^*)}} = c^* . \tag{3}$$

In the next section we consider the special cases of distributions of ability to process information such that:

$$F(c) = \frac{c^m}{K_m} \quad , \; m > 0 \, , \; K_m > 0 \, .$$

(For example if $m = 1$, this is the uniform distribution on $[0, K_1]$. More generally, the m^{th} distribution is concentrated on $[0,B_m]$ where $B_m = ((m+1)\,K_m)^{\frac{1}{m+1}}$.)

$$\text{Then}: \quad c^* = \left[\frac{r}{2\sqrt{2\pi}}\right]^{\frac{2}{2+m}} K_m^{\frac{1}{2+m}} \, n^{-\frac{m}{2+m}}$$

IV. OPTIMAL SAMPLE SIZE

The role of the decision maker is to determine the sample size, on the basis of his prior information, to maximize the presample expected valuation of the project minus the induced private informational costs.

We assume that the decision maker approximates the distribution of evaluations in the population of size N by a normal distribution $N(\nu,\sigma^2)$ with an unknown mean ν on which he has a prior $N(0,p^2)$ such that $\sigma^2 + p^2 = r^2$, to ensure that the decision maker has the same information as any other agent.

The decision maker knows that the agents who do not buy information answer zero.

Let $\xi = \dfrac{\sum_{i=1}^{n_2} v_i}{n_2}$, the mean of the answers of the agents who buy information[7]. The posterior mean in the subsample of size n_2 is ξ. The posterior mean for the entire population is

$$\frac{n_2 \, p^2 \, \xi}{\sigma^2 + n_2 \, p^2} \, .$$

[7] We assume that the event, an agent has a true $v_i = 0$, whose probability is zero, does not occur.

Therefore, for a decision maker who is assumed to be risk neutral, the *post sample* evaluation of the project is:

$$(N - n_2)\ \left(\frac{n_2 p^2\,\xi}{\sigma^2 + n_2 p^2}\right) + n_2\,\xi \;=\; \frac{(Np^2 + \sigma^2)\,\xi}{\frac{\sigma^2}{n_2} + p^2}\,.$$

Let h (ξ) be the *ex ante* distribution of ξ ; it is clearly a normal distribution of mean zero and variance $\frac{\sigma^2}{n_2} + p^2$.

The decision maker accepts the project if the sum of the answers is positive: $\xi \geqslant 0$.[8] Therefore the pre-sample evaluation of the project is:

$$V\,(N,n) \;=\; \int_{\xi \geqslant 0} \frac{(N\,p^2 + \sigma^2)}{\left(\frac{\sigma^2}{n_2} + p^2\right)}\ \xi\ e^{-\frac{1}{2}\frac{\xi^2}{\left(\frac{\sigma^2}{n_2} + p^2\right)}}\ d\xi$$

$$=\quad \frac{N\,p^2 + \sigma^2}{\sqrt{2\pi}\sqrt{\frac{\sigma^2}{n_2} + p^2}} \qquad \text{with}\quad n_2 = n\,F\,(c^*\,(n))\,.$$

The expected private informational costs are:

$$\int_0^{c^*(n)} c\ d\,F\,(c)\,.$$

The net value of the experiment with a sample size n is:

$$G\,(n\,,N) \;=\; \frac{N\,p^2 + \sigma^2}{\sqrt{2\pi}\sqrt{\frac{\sigma^2}{n\,F\,(c^*(n))} + p^2}} - \int_0^{c^*(n)} c\ d\,F\,(c)\,.$$

The optimal n is a solution of

$$\frac{\partial V}{\partial n}\,(n\,,N) \;=\; 0 \quad \text{if} \quad n > 0 \text{ and } n < N.$$

To gain some intuition regarding this solution we use the approximation developed at the end of section II.

[8] Note that here the decision would be identical if it were based on the mean of the posterior distribution.

$$\text{Since } F(c) = \frac{c^m}{K_m} \quad \text{for} \quad 0 \leqslant c \leqslant K_m^{\frac{1}{m}},$$

$$n \int_0^{c^*(n)} c \, d\, F(c) = \frac{n\,m}{m+1} \cdot \frac{c^{m+1}(n)}{K_m}.$$

The presample evaluation of the project is then:

$$G(n, N) = \frac{N\,p^2 + \sigma^2}{\sqrt{2\pi}\,\sqrt{p^2 + \frac{8\pi}{r^2}\sigma^2 c^{*2}(n)}} - \frac{n\,m}{m+1} \cdot \frac{c^*(n)^{m+1}}{K_m}$$

$$\text{with} \quad c^*(n) = \left[\frac{r}{2\sqrt{2\pi}}\right]^{\frac{2}{2+m}} K_m^{\frac{1}{2+m}} \; n^{-\frac{m}{2+m}}.$$

Differentiating with respect to n and equating to zero we obtain the optimal sample size. For m = 1 (uniform distribution) this is:

$$n \approx \frac{8\sqrt{2\pi}\;K_1\,\sigma^2}{pr^2} \cdot N.$$

More generally, we can ascertain the rate of growth of the optimal sample size for $m < 1$ by differentiating G (n, N) with respect to n and setting $n = N^{\alpha}$. The value of α for which this derivative is zero asymptotically is $\alpha = \frac{2+m}{2m+2-m^2}$. For example if m = ½, n* (N) is proportional, in limit, to $N^{10/11}$ and if m = 1/10, n* (N) is proportional, in the limit, to $N^{210/219}$.[9]

[9] The coefficient of proportionality is given by

$$\frac{m+1}{2-m^2} \cdot \frac{8\pi\sigma^2}{r^2 p} \left(\frac{r}{2\sqrt{2\pi}}\right)^{\frac{2}{2+m}} K_m^{\frac{1-m}{2+m}} \cdot N$$

V. INFORMATIONAL VALUE

In the previous sections we have deducted that the optimal sample size should grow at the rate of $N^{(2+m)/(2+2m-m^2)}$, where N is the population size and $0 \leqslant m \leqslant 1$ is the parameter of the structure of private information costs as we have specified it. In this way the three conflicting problems of inaccurate answers, sampling variance and private information gathering costs are traded off to best advantage.

Because the process involves real resource costs, however, it is necessary to demonstrate that the procedure is of positive value to society. More specifically, as N is large, we want to ascertain the asymptotic rate of growth of the net value of the procedure. A genuine informational advantage of this method can be said to follow if this value is positive and if its limit is positive when divided by N. As we increase the population size in our conceptual experiment, we are implicitly increasing the scale of the public project under consideration—therefore the *per capita* informational value is a sign of the actual informational value of the mechanism in the presence of inaccurate, costly, personal information regarding preferences.

Substituting into G, $n = N^{(2+m)/(2+2m-m^2)}$ we obtain only a lower bound on the net gain, since the true optimal sample size is on this order but may differ from it by lower order terms. This gives:

$$G(n^*,N) \geqslant \frac{Np^2 + \sigma^2}{\sqrt{2\pi}\sqrt{p^2 + \frac{8\pi A^2\sigma^2}{r^2} N^{(-2m)/(2+2m-m^2)}}}$$

$$- \frac{m}{m+1} \frac{A^{m+1}}{K_m} N^{(2-m^2)/(2+2m-m^2)}$$

where

$$A = \left(\frac{r}{2\sqrt{2\pi}}\right)^{\frac{2}{2+m}} K_m^{\frac{1}{2+m}} .$$

The first term is positive and its denominator is decreasing in N, approaching $p\sqrt{2\pi}$. The second, negative term is growing at the rate

$\frac{2-m^2}{2+2m-m^2}$ which is less than one for $0 \leqslant m \leqslant 1$.

Therefore $G(n^*,N)$ is asymptotically positive and, moreover,

$$\lim_{N\to\infty} \frac{G(n^*,N)}{N} = \frac{p}{\sqrt{2\pi}} > 0 .$$

VI. CONCLUDING REMARKS

We can now summarize the working of the model as follows. When the sample size n increases, the proportion of agents buying information decreases since $c^*(n)$ decreases. However, the absolute number of agents buying information increases since $n\ F\ (c^*(n)) \approx n^{\frac{2m}{2+m}}$. Since in this model the agents who do not buy information do not introduce noise in the answers (they are easily recognized since they answer 0 and hence are effectively deleted from the sample), it is clear that without informational costs the optimal sample size would be N. However private informational costs increase as $n^{\frac{2-m^2}{2+m}}$, so that there is an optimal size of the order of $N^{\frac{2+m}{2m+2-m^2}}$.

If the answers of the agents not buying information were not detectable, this noise could well lead to a decreasing number of agents buying information after a given sample size, so that the optimal sample size would be essentially bounded (see Green and Laffont [1975] for such a model). Then in a way we could say that the Clarke mechanism is not really applicable since only a very small proportion of the population is really sampled, leading to a large potential sampling error. In the model presented here this does not happen for the reasons explained above. There may still be a difficulty if there is a correlation between the v_i and the c_i, i.e. between the evaluations of the public good and the ability to process information, since we use only answers given by agents with low c_i. The conditional expectation of v given that c is larger than a cut off point c^* may be different from the unconditional expectation in which we are interested. It is not easy to correct the induced bias. If we want to draw a subsample to estimate the correlation (by paying everybody his information costs) and then use the estimation to suppress the bias in the large sample, we face the difficulty that in the subsample, truth revelation will not be in general a dominant strategy anymore. The costly solution of paying everybody in the random sample faces another type of difficulty. Information costs can be divided into information-gathering costs and information-processing costs. It is clear that the decision maker can pay the information-gathering costs, or, even better, use the likely increasing returns to scale of information dissemination. No increasing returns to scale seem to exist in the activity of information processing, and more importantly there is not way to be sure that, even when they are paid, agents will incur the (often psychological) costs of information processing.

REFERENCES

Bohm, Peter. "Estimating Demand for Public Goods: an Experiment." *Europ. Econ. Rev.,* 3 (1972), 111-130.

Bowen, Howard. "The Interpretation of Voting in the Allocation of Economic Resources." *Quart. Jour. Econ.,* 58 (1943), 27-48.

Clarke, E.H. "Multipart Pricing of Public Goods." *Public Choice*, 8 (1971), 19-33.

Green, Jerry, and Laffont, Jean-Jacques. "Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods." *Harvard Institute of Economic Research*, DP. N° 412 (1975a), to appear in *Econometrica.*

Green, Jerry, and Laffont, Jean-Jacques. "Strength of Incentives for S.I.I.C. Mechanisms." *Harvard Institute of Economic Research*, DP. N° 411 (1975b).

Green, Jerry, Laffont, Jean-Jacques. "Revelation des Preferences pour les Biens Publics: lere Partie, Caracterisation des Mecanismes Satisfaisants." *Ecole Polytechnique*, DP. A125 0176 (1976).

Groves, Theodore. "Incentives in Teams." *Econometrica*, 41 (1973), 617-631.

Kurz, Mordecai. "Experimental Approach to the Determination of the Demand for Public Goods." *Jour. Pub. Econ.*, 3 (1974), 329-348.

Musgrave, Richard. *The Theory of Public Finance*. New York, McGraw Hill, 1959.

Samuelson, Paul. "The Pure Theory of Public Expenditures." *Rev. Econ. Stat.*, 36 (1954), 387-389.

Thompson, E. A.. "A Pareto Optimal Group Decision Process." *Papers on Non-Market Decision Making*, University of Virginia, 1967.

Wicksell, Kurt. *Finanztheoretische Untersuchungen und das Steuerwesen* Sweden's, Jena, Germany, 1896.