Regional externalities and efficient decentralization under incomplete information*

Claude d'Aspremont
CORE, Université catholique de Louvain
Louis-André Gérard-Varet
Université Louis Pasteur, Strasbourg I and G.R.E.Q.E. - Era CNRS983, Universités Aix-Marseille.

Abstract

In an interregional context we formalize the problem of individual incentives and collective efficiency as an externality game with incomplete information. More specifically, the problem is to find a cost-redistribution scheme which implies incentive compatibility in the sense of a Bayesian Equilibrium. Two classes of such redistribution schemes are defined. In the case of independent beliefs, and using differentiability assumptions, it is shown that one of these two classes coincide with the class of all incentive compatible schemes and that the other contains schemes which are incentive compatible in a stronger sense. Finally, we construct an incentive compatible scheme with no budgetary consequences.

1. Introduction

The purpose of this paper is to point out an informational problem which arises in the context of international or interregional relationships arrangements. One example, which is treated here, is when one region is involved in an economic activity which has external effects (beneficial of damaging) on some other regions and whenever a supra-regional agency has to elaborate a compensatory scheme preserving collective efficiency. Another example is when a country divided into regions, some supra-regional agency has to determine the production of a a public good, or service, at a level which should be efficient in terms of the benefits accruing to the regions, as they are assessed by the regions themselves. In both cases, the supra-regional agency must base its decision upon informations communicated by the regions. Thus, it may well be the case that the regions find advantageous to distort the information they communicate to the agency.

Some basic features of the problem may be captured in the framework of a simple situation involving two regions where some activity of the first region creates a certain level of a negative externality which affects the second region. This negative externality may be, for example, the level of pollution dumped by the first region in a river crossing the second region, or the nuisance generated by an obnoxious facility located between population centers of two different countries. The central agency tries to regulate this level according to an efficiency criterium and to share the

*. This paper is a revised version of CORE Discussion Paper no. 7519. Some of the results presented here have been generalized in subsequent papers (see d'Aspremont and Gérard-Varet, 1979a,b). However some other results are linked to the particular approach we adopt here and which was first used by Smets (1973) and later on by Laffont and Maskin (1979).
Reprinted from *Locational Analysis of Public Facilities*, J.-F. Thisse and H.G. Zoller (eds.), North-Holland Publishing Company, 1983.

resulting costs, on the basis of the evaluation provided by the two regions. This particular example has been studied first by Smets (1973). However, the cost-redistribution scheme suggested[1], although it has an interesting incentive property, does not eliminate an important budgetary problem: the central agency may keep a surplus or a deficit. We want to point out here the interest of formulating the problem as a non-cooperative game with incomplete information, and thus of taking into account the "beliefs" of the participants.

To be more specific we consider a situation involving two regions and a central agency which has to fix the level of a pollution produced in one region and affecting the other. Each region is described by a certain number of characteristics and alone knows their true values. For every level of pollution, we associate a cost for the first region, representing the associated purification expenses, and a cost for the second region, representing the damages incurred. The cost function of each country is parametrized by its own characteristics. In addition a scheme for redistributing the costs is given, which is not a function of the true characteristics. Finally, each region is supposed to declare some plausible value for its characteristics vector, to have some "beliefs" concerning the other region's true characteristics and to know both the cost-redistribution scheme and the rule according to which the agency fixes pollution at the level minimizing the sum of the "declared" costs. In the general case, the beliefs of each region may depend on its own true characteristics. However, some of our results will assume independence.

In terms of this model, the incentive problem is to find a cost-redistribution scheme which is incentive compatible in the sense that each region will *believe* it to be in its self-interest to declare its true characteristics if it knows that the other is doing the same. Moreover, such a scheme whose operation results in selection of the optimal level of pollution is called collectively efficient if, in addition, the budget of the agency remains balanced whatever the declarations of the two regions. In Section 2, the model is given formally.

Section 3 presents two classes of redistribution schemes which are shown to be incentive compatible. Also, it is shown that in the case where the beliefs are independent, the second of these classes includes all incentive compatible schemes and that the schemes of the first class are incentive compatible in a stronger sense. In Section 4, we demonstrate that, in the independence case, there are some redistribution schemes which not only are incentive compatible but also allow the agency to keep its budget balanced. However, such schemes may not belong to the first class i.e., they may not enjoy incentive compatibility in the stronger sense.

2. The model

2.1. In the context we have just defined we call the polluting region player I and the polluted region player II. Now, following Harsanyi (1967-68), we assume that each player is described by a number of characteristics. More specifically we assume here that the characterization of player I is determined by a parameter α belonging to a characteristic subspace $\mathcal{A}$ of $\mathbb{R}^n$ and that the characterization of player II is determined by a multi-valued parameter β belonging to a characteristic subspace $\mathcal{B}$ of $\mathbb{R}^n$. When the true value of the characteristics of player I (resp. player II) is α (resp. β) we then say that player I (resp. player II) is of type α (resp. type β). Hence $\mathcal{A}$ is the space of all possible types for player I and $\mathcal{B}$ the space of all possible types for player II. Now to each player of each type we associate a nonnegative cost function which is a real-valued function of a nonnegative real

1. There exists now a considerable amount of literature on this kind of scheme. For references see the book by Green and Laffont (1979).

variable p denoting the level of pollution. For player I of type α this cost function is the abatement cost function $C(p,\alpha)$ and for player II of type β it is the damage cost function $D(p,\beta)$. The general form of the functions C and D is supposed to be known by the two players and the agency. However, player I alone knows the true value of the parameter α and player II alone knows the true value of the parameter β. The general form of C and D is restricted as follows:

c_1: $\mathcal{A}$ is a bounded open subset of $\mathbb{R}^n$.

c_2: $C(p,\alpha)$ is a twice continuously differentiable function in $\mathbb{R}_+ \times \mathbb{R}^n$.

c_3: $C(p,\alpha)$ is positive and strictly decreasing in p for any α, i.e., $\forall\,\alpha \in \mathbb{R}^n, \forall\, p \in \mathbb{R}_+ C(p,\alpha) > 0$, $C_p(p,\alpha) \stackrel{\text{def}}{=} \frac{\partial}{\partial p} C(p,\alpha) < 0$.

c_4: $C(p,\alpha)$ is strictly convex in p for any α, i.e., $\forall\,\alpha \in \mathbb{R}^n, \forall\, p \in \mathbb{R}_+ C_{pp}(p,\alpha) \stackrel{\text{def}}{=} \frac{\partial^2}{\partial p^2} C(p,\alpha) > 0$.

c_5: $\forall\, p \in \mathbb{R}_+, \forall\,\alpha \in \mathbb{R}^n, C_{p\alpha}(p,\alpha) \stackrel{\text{def}}{=} \left(\frac{\partial}{\partial \alpha_1} C_p(p,\alpha), \ldots, \frac{\partial}{\partial \alpha_n} C_p(p,\alpha)\right) \neq 0$.

d_1: $\mathcal{B}$ is a bounded open subset of $\mathbb{R}^n$.

d_2: $D(p,\beta)$ is a nonnegative bounded twice continuously differentiable function in $\mathbb{R}_+ \times \mathbb{R}^n$.

d_3: $\forall\,\beta \in \mathbb{R}^n, \forall\, p \in \mathbb{R}_+, D_p(p,\beta) \stackrel{\text{def}}{=} \frac{\partial}{\partial p} D(p,\beta) > 0$.

d_4: $\forall\,\alpha \in \mathbb{R}^n, \forall\,\beta \in \mathbb{R}^n, \forall\, p \in \mathbb{R}_+; D_{pp}(p,\beta) \stackrel{\text{def}}{=} \frac{\partial}{\partial p} D_p(p,\beta) > -C_{pp}(p,\alpha)$.

d_5: $\forall\, p \in \mathbb{R}_+, \forall\,\beta \in \mathbb{R}^n, D_{p\beta}(p,\beta) \stackrel{\text{def}}{=} \left(\frac{\partial}{\partial \beta_1} D_p(p,\beta), \ldots, \frac{\partial}{\partial \beta_n} D_p(p,\beta)\right) \neq 0$.

d_6: $\forall\,\alpha \in \mathbb{R}^n, \forall\,\beta \in \mathbb{R}^n, \exists\, p \in (0,\infty)$ such that $C_p(p,\alpha) + D_p(p,\beta) = 0$.

The assumptions c_1–c_2 and d_1–d_2 are for technical convenience. Assumptions c_3 and d_3 are easily interpreted. The asymmetry between assumptions c_4 and d_4 is introduced because it seems more natural to impose convexity restrictions on C than on D. Assumptions c_5 and d_5 characterize rather strongly the impact of the parameters α and β. Finally d_6 ensures the existence of a critical point in $(0,\infty)$ for the functions $C(p,\alpha) + D(p,\beta)$. This assumption may be justified, for example, by assuming that when pollution increases, the decrease in purifying costs is higher near the zero point than the increase in damages, while the reverse holds at some high pollution level.

2.2. In the same way that we have assumed that we can associate to each player of each type a particular cost function, we now assume that each player of each type possesses some kind of "beliefs" concerning the type of the other player.

Formally, we introduce, for player I and each type α, a subjective probability density function $\mu(\cdot \mid \alpha)$ on $\mathcal{B}$. Similarly, for player II and each type β, we introduce a subjective probability density function $\nu(\cdot \mid \beta)$ on $\mathcal{A}$. Both μ and ν are defined on all $\mathbb{R}^n \times \mathbb{R}^n$.

Even if the agency is supposed not to know the type of each player, we assume that it knows the space of all possible types and the set of all possible subjective probability density functions for each player.

2.3. For every value of the pair (α, β) denoting a pair of true types, we shall now associate a particular non-cooperative bimatrix game $\Gamma_{(\alpha,\beta)}$ by defining first the respective *strategy spaces* A_α and B_β of players I and II and then their respective *payoff functions* V and W.

For each game $\Gamma_{(\alpha,\beta)}$ a strategy of player I will consist of a certain value $a \in \mathbb{R}^n$, denoting the *declared* value of his true parameter α, and a strategy of player II will consist of a certain value $b \in \mathbb{R}^n$, denoting the *declared* value of his true parameter β. Formally this means that, for each pair (α, β), A_α and B_β are subspaces of $\mathbb{R}^n$. We shall actually go further by assuming that the strategy space of both players can be identified to their parameter spaces, i.e., there exist open bounded subsets A and B of $\mathbb{R}^n$ such that[2]:

$$\forall\, \alpha \in \mathcal{A}, \forall\, \beta \in \mathcal{B}, A_\alpha = \mathcal{A} = A \text{ and } B_\beta = \mathcal{B} = B.$$

Since a strategy for a player consists of announcing a type which the agency (or the other player) must believe to be his true type, it seems natural to assume that each possible type can be announced and no other. However, to clarify the presentation we shall use small greek letters to denote true types and small latin letters to denote declared types.

Given a pair of strategies $(a, b) \in A \times B$, the agency determines the respective payoffs $V(a, b; \alpha)$ and $W(a, b; \beta)$ of players I and II according to the following rules:

r_1: *collective efficiency rule*: the agency fixes pollution at the level $p(a, b)$ which minimizes the sum of the *declared costs*, namely $C(p, a) + D(p, b)$.

Note that by assumptions $c_2 - c_4$, $d_2 - d_4$ and d_6, $p(a, b)$ is uniquely determined by the following condition:

$$p(a, b) = p \text{ if and only if } C_p(p, a) + D_p(p, b) = 0.$$

By the implicit function theorem[3] $p(a, b)$ is a continuously differentiable function on $A \times B$.

The next two rules taken together define what has been called a cost-redistribution scheme. The distinction is introduced for convenience.

r_2: *cost-sharing rule:* the agency receives in payment from player I the declared cost $D(p(a, b), b)$ of player II and from player II the declared cost $C(p(a, b), a)$ of player I.

r_3: *distribution rule:* the agency pays to player I a certain amount $F_I(a, b)$ and to player II a certain amount $F_{II}(a, b)$ according to a scheme fixed by the agency and the players independently of the true values α and β of the parameters. Functions F_I and F_{II} are supposed to be continuously differentiable real-valued functions on $\mathbb{R}^n \times \mathbb{R}^n$. By extension a pair of such functions (F_I, F_{II}) will be called a *distribution rule.*

2. We denote by $\overline{A}$ the closure of A and by $\overline{B}$ the closure of B.

3. (See Fleming, 1965, p. 117). By c_2, d_2 and d_4 the transformation f from $A \times B \times (0, \infty)$ to itself defined by:

$$\text{for } i = 1, 2, \ldots, n,\ f_i(a, b, p) = a_i,\ f_{n+i}(a, b, p) = b_i \text{ and } f_{2n+1}(a, b, p) = C_p(p, a) + D_p(p, b),$$

is a diffeormorphism of class $\mathcal{C}^{(1)}$. Hence f has an inverse g which is also of class $\mathcal{C}^{(1)}$. It is clear, by d_6, that $\{(a, b) \in A \times B : (a, b, 0) \in f(A \times B \times (0, \infty))\} = A \times B$, which is open. Letting $p(a, b) = g_{2n+1}(a, b, 0)$, the result follows.

Now according to these rules, the payoff functions for any true values α and β are:

$$V(a,b;\alpha) \stackrel{\text{def}}{=} C(p(a,b),\alpha) + D(p(a,b),b) - F_I(a,b), \text{ for player } I$$
$$W(a,b;\beta) \stackrel{\text{def}}{=} D(p(a,b),\beta) + C(p(a,b),a) - F_{II}(a,b) \text{ for player } II.$$

For each player the first amount is his actual cost associated to the pollution level $p(a,b)$, the second amount is the declared cost of the other player, also at the level $p(a,b)$, and the third amount is the payment made by the agency.

Remark *The functions $V(\cdot,\cdot;\alpha)$ and $W(\cdot,\cdot;\beta)$ are continuously differentiable on $\mathbb{R}^n \times \mathbb{R}^n$. Consequently, we can write for player I (for example)*

$$V_a(a,b;\alpha) = \frac{\partial}{\partial a} p(a,b)[C_p(p(a,b),\alpha) + D_p(p(a,b),b)] - \frac{\partial}{\partial a} F_I(a,b),$$

where

$$\frac{\partial}{\partial a} p(a,b) = \frac{-C_{pa}(p(a,b),a)}{C_{pp}(p(a,b),a) + D_{pp}(p(a,b)),b)}$$

which is different from zero by conditions c_4, d_4 and c_5.

2.4. We have thus defined a *non-cooperative game with incomplete information* Γ, in the sense of Harsanyi (1967-68), which is the set of all possible bimatrix games $\Gamma_{(\alpha,\beta)}$ together with the set of all possible pairs of density functions $\langle \mu(\cdot \mid \alpha), \nu(\cdot \mid \beta)\rangle$. We want now to introduce the concept of equilibrium which will appear to be the relevant solution in the non-cooperative and incomplete information context of the game Γ. For this purpose, following Harsanyi's terminology, we give a new description of the game in which the incomplete information aspects are integrated into new strategy spaces and new payoffs functions.

Let us first define new strategy spaces. *A normalized strategy* is a decision rule specifying for a player the set of strategies he would use in the game Γ, one for each given true value of his characteristics. Formally we shall assume that a normalized strategy for player I is a measurable function a^* from $\mathcal{A}$ into itself and a normalized strategy for player II is a measurable function b^* from $\mathcal{B}$ into itself. We denote respectively by A^* and B^* the spaces of all possible normalized strategies.

Now for each player of each type, we shall suppose that his actual objective function depends on his choice of a declared type and on the normalized strategy chosen by the other player and equals the expected payoff knowing his own true type. Hence the respective expected payoffs for player I of type $\alpha \in A$ and player II of type $\beta \in B$ are:

$$\forall\, a \in A,\ \forall\, b^* \in B^*, \overline{V}(a,b^*;\alpha) \stackrel{\text{def}}{=} \int_B V(a,b^*(\beta);\alpha)\mu(\beta \mid \alpha)d\beta$$
$$\forall\, b \in B,\ \forall\, a^* \in A^*, \overline{W}(a^*,b;\beta) \stackrel{\text{def}}{=} \int_A W(a^*(\alpha),b;\beta)\nu(\alpha \mid \beta)d\alpha.$$

By the definition of a^* and b^* and by the continuity of each $V(\cdot,\cdot;\alpha)$ and each $W(\cdot,\cdot;\beta)$ on $\mathbb{R}^n \times \mathbb{R}^n$, we see that $\overline{V}(\cdot,\cdot;\alpha)$ and $\overline{W}(\cdot,\cdot;\beta)$ are well defined functions on $A \times B^*$ and $A^* \times B$ respectively.

Finally, we get the following notion of equilibrium. We shall say that a pair $\langle a^*, b^*\rangle$ of normalized strategies forms a *Bayesian Equilibrium Pair* (BEP) if and only if $a^*(\alpha)$ minimizes $\overline{V}(a,b^*;\alpha)$ for every $\alpha \in A$ and $b^*(\beta)$ minimizes $\overline{W}(a^*,b;\beta)$ for every $\beta \in B$.

We note that although the actual objective function of each player is conditional on his true type, the associated notion of equilibrium must be defined with respect to every possible pair of types, since, a priori, any pair in $A \times B$ is admissible as a pair of true types.

3. Distribution rules and incentives

3.1. In the model we have just described, the individual incentive problem is whether some region of some type will find in its interest not to reveal its true type, whatever it may be. We shall say that we have a solution to this particular instance of the individual incentive problem whenever there exists a distribution rule such that revealing his true type, for each player and each type, forms a BEP in the game Γ associated to this particular rule. Hence we shall give the name *Incentive Compatible Distribution Rule* (ICDR) to any distribution rule (F_I, F_{II}) such that, for the associated game Γ, the pair of normalized strategies $\langle \hat{a}^*, \hat{b}^* \rangle$ defined by

$$\forall\, \alpha \in A, \hat{a}^*(\alpha) = \alpha,$$
$$\forall\, \beta \in B, \hat{b}^*(\beta) = \beta,$$

is a BEP in Γ.

In this section we want not only to characterize a class of ICDR's but also, in this class, to isolate a subclass of distribution rules having in addition a *uniqueness* property. For this purpose, we define a stronger rule namely: a *Strongly Incentive Compatible Distribution Rule* (SICDR) is an ICDR such that, for any BEP $\langle a^*, b^* \rangle$ in the associated game Γ,

$$p(a^*(\alpha), b^*(\beta)) = p(\hat{a}^*(\alpha), \hat{b}^*(\beta))$$

a.e., with respect to the Lebesgue product measure denoted $\lambda_1 \times \lambda_2$. This means essentially that if, for some reason, a BEP other than $\langle \hat{a}^*, \hat{b}^* \rangle$ is reached, the "truly" collectively optimal level of pollution will still be imposed by the agency.

3.2. In fact, we shall limit our investigation to the following two types of distribution rules.

1. A distribution rule (F_I, F_{II}) is a *Discretionary Distribution Rule* (DDR) if $\forall\, a \in A, \forall\, b \in B$,

   (a) $\forall a' \in A, F_I(a', b) = F_I(a, b)$,
   (b) $\forall b' \in B, F_{II}(a, b') = F_{II}(a, b)$.

   The term "discretionary" in the definition is to reflect the fact that F_I must be constant in A and F_{II} constant in B.

2. A distribution rule (F_I, F_{II}) is a *Subjectively Discretionary Distribution Rule* (SDDR) if $\forall\, a \in A, \forall\, a' \in A, \forall\, b \in B, \forall\, b' \in B$,

   (a) $\forall \alpha \in A, \displaystyle\int_B F_I(a', \beta)\mu(\beta \mid \alpha) d\beta = \int_B F_I(a, \beta)\mu(\beta \mid \alpha) d\beta$
   (b) $\forall \beta \in B, \displaystyle\int_A F_{II}(\alpha, b')\nu(\alpha \mid \beta) d\alpha = \int_A F_{II}(\alpha, b)\nu(\alpha \mid \beta) d\alpha.$

It is easy to see that any DDR is a SDDR. Hence the term "subjectively discretionary" reflects the fact that only the expected conditional value of F_I must be constant in A and the expected conditional value of F_{II} must be constant in B. Note also that to fix such a rule the agency must use its information about all the possible conditional subjective density functions $\mu(\cdot \mid \alpha)$ and $\nu(\cdot \mid \beta)$.

3.3. We may now present some results describing the relation between ICDR's and SDDR's. The first proposition shows that any SDDR is an ICDR. Now, since any DDR is a SDDR, it is clear that the class of all SDDR's is non-empty and so the first proposition implies the existence of an ICDR. The second proposition shows that in addition any ICDR is a SDDR in the case where the beliefs of each player are independent of their own types; namely whenever

$$\forall \alpha \in A, \forall \alpha' \in A, \forall \beta \in B, \mu(\beta) \stackrel{\text{def}}{=} \mu(\beta \mid \alpha) = \mu(\beta \mid \alpha') \tag{1}$$

$$\forall \beta \in B, \forall \beta' \in B, \forall \alpha \in A, \nu(\alpha) \stackrel{\text{def}}{=} \nu(\alpha \mid \beta) = \nu(\alpha \mid \beta'). \tag{2}$$

Hence we obtain a characterization of ICDR's. Finally the third proposition shows that any DDR is not only an ICDR but also a SICDR.

Proposition 1 *The class of all SDDR's is contained in the class of all ICDR's.*

Proof Let (F_I, F_{II}) be a SDDR. We shall show that

(i) $\forall \alpha \in A,\ \forall a \in A, \overline{V}(\alpha, \hat{b}^*; \alpha) \leq \overline{V}(a, \hat{b}^*; \alpha)$,

(ii) $\forall \beta \in B,\ \forall b \in B, \overline{W}(\hat{a}^*, \beta; \beta) \leq \overline{W}(\hat{a}^*, b; \beta)$.

Let us treat case (i) only. Case (ii) follows by symmetry. By rule r_1 we know

(iii) $\forall \beta \in B, \forall \alpha \in A,\ \forall a \in A, C(p(\alpha, \beta), \alpha) + D(p(\alpha, \beta), \beta) \leq C(p(a, \beta), \alpha) + D(p(a, \beta), \beta)$.

It is easy to see that (iii) implies:

(iv) $\forall \alpha \in A,\ \forall a \in A, \int_B [C(p(\alpha, \beta), \alpha) + D(p(\alpha, \beta), \beta)]\mu(\beta \mid \alpha) d\beta$
$\leq \int_B [C(p(a, \beta), \alpha) + D(p(a, \beta), \beta)]\mu(\beta \mid \alpha) d\beta.$

But, by the fact that (F_I, F_{II}) is a SDDR we have

(v) $\forall \alpha \in A,\ \forall a \in A, \int_B F_I(\alpha, \beta)\mu(\beta \mid \alpha) d\beta = \int_B F_I(a, \beta)\mu(\beta \mid \alpha) d\beta$.

The result follows since (v) implies that (iv) is equivalent to (i). ■

Proposition 2 *In the case where (1) and (2) hold, the class of all ICDR's coincides with the class of all SDDR's.*

Proof By Proposition 1, we know that any SDDR is an ICDR. It remains to show that any ICDR is a SDDR. Let (F_I, F_{II}) be some ICDR. We shall show that

(i) $\forall a \in A, \dfrac{\partial}{\partial a} \int_B F_I(a, \beta)\mu(\beta) d\beta = 0$

(ii) $\forall b \in B, \dfrac{\partial}{\partial b} \int_A F_{II}(\alpha, b)\nu(\alpha) d\alpha = 0,$

which, by our assumptions, is equivalent to the defining condition of a SDDR. We shall derive only (i) since (ii) follows by symmetry.

Since (F_I, F_{II}) is an ICDR, $\langle \hat{a}^*, \hat{b}^* \rangle$ is a BEP. Also for any type α of player I, $\overline{V}(\cdot, \hat{b}^*; \alpha)$ is differentiable on A. Indeed we know that for any type α:

$$\forall\, a \in \mathbb{R}^n, \forall\, \beta \in \mathbb{R}^n, V_a(a, \beta; \alpha) = \frac{\partial}{\partial a}[C(p(a, \beta), \alpha) + D(p(a, \beta), \beta) - F_I(a, \beta)]$$

is continuous (see remark, p. 7). So in particular any function $V_a(\cdot, \cdot; \alpha)$ is continuous on the cartesian product of the open set A by the closure $\overline{B}$ of B (which is compact). Also, by the definition of μ, we may write:

$$\forall\, a \in A, \frac{\partial}{\partial a}\int_B V(a, \beta; \alpha)\mu(\beta)d\beta = \frac{\partial}{\partial a}\int_B V(a, \beta; \alpha)\mu(\beta)d\beta.$$

Therefore we get (see e.g. Fleming, 1965, p. 199):

$$\forall\, a \in A, \overline{V}_a(a, \hat{b}^*; \alpha) = \frac{\partial}{\partial a}\int_B V(a, \beta; \alpha)\mu(\beta)d\beta = \int_B V_a(a, \beta; \alpha)\mu(\beta)d\beta.$$

Now since this shows that, for any $\alpha \in A$, $\overline{V}(\cdot, \hat{b}^*; \alpha)$ is differentiable on the open set A and since, $\langle \hat{a}^*, \hat{b}^* \rangle$ being a BEP, $V(\hat{a}, \hat{b}^*; \alpha)$ is minimum for $a = \hat{a}^*(\alpha)$, we must have, for $a = \hat{a}^*(\alpha) = \alpha$,

$$\begin{aligned}\overline{V}_a(a, \hat{b}^*; \alpha) = & \int_B [\frac{\partial}{\partial a}p(a, \beta][C_p(p(a, \beta), \alpha) + D_p(p(a, \beta), \beta)]\mu(\beta)d\beta \\ & - \int_B \frac{\partial}{\partial a}F_I(a, \beta)\mu(\beta)d\beta = 0.\end{aligned}$$

But, by rule r_1, $[C_p(p(\alpha, \beta), \alpha) + D_p(p(\alpha, \beta), \beta)] = 0$ and hence,

$$\begin{aligned}\forall\, \alpha \in A, \overline{V}_a(\alpha, \hat{b}^*; \alpha) &= -\int_B \frac{\partial}{\partial a}F_I(\alpha, \beta)\mu(\beta)d\beta = -\int_B \frac{\partial}{\partial a}F_I(\alpha, \beta)\mu(\beta)d\beta \\ &= -\frac{\partial}{\partial a}\int_B F_I(a, \beta)\mu(\beta)d\beta = 0 \qquad a = \alpha.\end{aligned}$$

Finally, because the symmetric argument applies to player II, we see that (F_I, F_{II}) must be a SDDR. ■

For the next proposition we assume that:

$$\forall\, \alpha \in A, \forall\, \beta \in B, \mu(\beta \mid \alpha) > 0 \text{ and } \nu(\alpha \mid \beta) > 0. \tag{3}$$

Proposition 3 *If μ and ν satisfy (3), then the class of all DDR's is contained in the class of all SICDR's.*

Proof Let (F_I, F_{II}) be any DDR. By Proposition 1, we know that $\langle \hat{a}^*, \hat{b}^* \rangle$ is a BEP. Let $\langle \overline{a}^*, \overline{b}^* \rangle$ be any other BEP.

1. As a first step we shall show that:

   (i) $\forall\, \alpha \in A, p(\alpha, \overline{b}^*(\beta)) = p(\overline{a}^*(\alpha), \overline{b}^*(\beta))$ a.e. (with respect to λ_2), (i') $\forall\,\beta \in B, p(\overline{a}^*(\alpha),\beta) = p(\overline{a}^*(\alpha),\overline{b}^*(\beta))$ a.e (with respect to λ_1).

We shall only prove (i), since (i') follows by symmetry. First, for any $\alpha \in A$, let μ_α be a probability measure defined on $\mathbb{R}^n$ such that for any Borel set E of $\mathbb{R}^n$

$$\mu_\alpha(E) \stackrel{\text{def}}{=} \int_E \mu(\beta \mid \alpha)d\beta.$$

Then, by the fact $\langle \overline{a}^*, \overline{b}^* \rangle$ is a BEP and (F_I,F_{II}) a DDR, we have

(ii) $\forall\,\alpha \in A, \overline{a}^*(\alpha)$ minimizes $\int_B [C(p(\cdot,\overline{b}^*(\beta)),\alpha) + D(p(\cdot,\overline{b}^*(\beta)),\overline{b}^*(\beta))]d\mu_\alpha(\beta)$.

Moreover, for any $\alpha \in A$, let $\overline{\mu}_\alpha$ be the probability measure induced on $\mathbb{R}^n$ by $\overline{b}^*$ in the following manner: for any Borel set G of $\mathbb{R}^n$, let

$$\overline{\mu}_\alpha(G) \stackrel{\text{def}}{=} \mu_\alpha(\overline{b}^{*-1}(G)).$$

Accordingly, (ii) becomes (see Halmos, 1959, p. 103)

(iii) $\forall\,\alpha \in A, \overline{a}^*(\alpha)$ minimizes $\int_B C(p(\cdot,b),\alpha) + D(p(\cdot,b),b)d\overline{\mu}_\alpha(b)$.

Now we need the following additional notation. We denote by Π the set of continuous functions from B to $\mathbb{R}_+$. Clearly, for any $\alpha \in A$, $p(a,\cdot) \in \Pi$. Also, for every $a \in A$ and $\pi \in \Pi$, let $\forall\, b \in B, h(\pi(b),b;\alpha) \stackrel{\text{def}}{=} C(\pi(b),\alpha) + D(\pi(b),b)$, and $H(\pi;\alpha) \stackrel{\text{def}}{=} \int_B h(\pi(b),b;\alpha)d\overline{\mu}_\alpha(b)$. Then, by rule r_1, we have for any $\alpha \in A$,

$$\forall\,\pi \in \Pi, \forall\, b \in B, h(p(\alpha,b),b;\alpha) \leq h(\pi(b),b;\alpha)$$

and so

(iv) $\forall\,\pi \in \Pi, H(p(\alpha,\cdot);\alpha) \leq H(\pi;\alpha)$.

In particular we get

$$H(p(\alpha,\cdot);\alpha) \leq H(p(\overline{a}^*(\alpha),\cdot);\alpha).$$

On the other hand, (iii) gives $H(p(\overline{a}^*(\alpha),\cdot);\alpha) \leq H(p(\alpha,\cdot);\alpha)$. Hence, putting these derivations together, we get for any $\alpha \in A$

(v) $H(p(\overline{a}^*(\alpha),\cdot);\alpha) = H(p(\alpha,\cdot);\alpha)$.

Finally, we shall derive (i) by contradiction. Suppose that, in contradiction with (i), there exists a set $E_\alpha \subseteq B$ of positive Lebesgue measure, i.e. $\lambda_2(E_\alpha) > 0$, such that:

$$\forall\,\beta \in E_\alpha, p(\alpha,\overline{b}^*(\beta)) \neq p(\overline{a}^*(\alpha),\overline{b}^*(\beta)).$$

Then, by (3), we have also $\mu_\alpha(E_\alpha) > 0$ or $\overline{\mu}_\alpha(\overline{b}^*(E_\alpha)) > 0$. This assumption implies the existence of a compact subset K of $\overline{b}^*(E_\alpha)$ such that $\overline{\mu}_\alpha(K) > 0$ (see Halmos, 1959). But by the strict convexity of $[C(\cdot,\alpha) + D(\cdot,b)]$ in p for every $b \in B$, we get

(vi) $\forall\, b \in B, \forall\,\lambda \in (0,1),\ h(\lambda p(\alpha,b) + (1-\lambda)p(\overline{a}^*(\alpha),b),b;\alpha)$
$\leq \lambda h(p(\alpha,b),b;\alpha) + (1-\lambda)h(p(\overline{a}^*(\alpha),b),b;\alpha)$,

with strict inequality holding if $b \in K$. The continuity of $h(\cdot;\alpha)$ on B and the compactness of K imply the existence of a positive number k such that

(vii) $k = \min_{b\in K}[\lambda h(p(\alpha,b),b;\alpha) + (1-\lambda)h(p(\overline{a}^*(\alpha),b),b;\alpha).$
$\quad - h(\lambda p(\alpha,b) + (1-\lambda)p(\overline{a}^*(\alpha),b),b;\alpha]$

Hence, by (v), (vi) and (vii),

(viii) $\quad H(\lambda p(\alpha,\cdot) + (1-\lambda)p(\overline{a}^*(\alpha),\cdot);\alpha) + k$
$\quad \leq \lambda H(p(\alpha,\cdot);\alpha) + (1-\lambda)H(p(a^*(\alpha),\cdot);\alpha) = H(p(\alpha,\cdot);\alpha),$

which contradicts (iv). Since, for player II, a similar contradiction may be derived for (i'), the first step of the proof is completed.

2. Suppose (i) and (i') hold.
We want to show that this implies

$$p(\overline{a}^*(\alpha),\overline{b}^*(\beta)) = p(\alpha,\beta) \text{ a.e. (with respect to } \lambda_1 \times \lambda_2).$$

Suppose, on the contrary, that there exists a subset $E \times F$ of $A \times B$ such that:

(ix) $\lambda_1(E)\cdot\lambda_2(F) > 0,$ and .
$\quad \forall\,(\alpha,\beta) \in E \times F, p(\overline{a}^*(\alpha),b^*(\beta)) \neq p(\alpha,\beta)$

By (i), $\lambda_2(F) > 0$ implies

(x) $\forall\,\alpha \in A, \forall\,\beta \in F, C_p(p(\overline{a}^*(\alpha),\overline{b}^*(\beta)),\alpha = -D_p(p(\overline{a}^*(\alpha),\overline{b}^*(\beta)),\overline{b}^*(\beta)).$

Similarly, $\lambda_1(E) > 0$ implies (by (i')):

(xi) $\forall\,\beta \in B, \forall\,\alpha \in E, C_p(p(\overline{a}^*(\alpha),\overline{b}^*(\beta)),\overline{a}^*(\alpha)) = -D_p(p(\overline{a}^*(\alpha),\overline{b}^*(\beta)),\beta).$

But, by rule r_1, page 6, we have:

(xii) $\forall\,\alpha \in A, \forall\,\beta \in B, C_p(p(\overline{a}^*(\alpha),\overline{b}^*(\beta)),\overline{a}^*(\alpha)) = -D_p(p(\overline{a}^*(\alpha),\overline{b}^*(\beta)),\overline{b}^*(\beta)),$

and

(xiii) $\forall\,\alpha \in A, \forall\,\beta \in B, p = p(\alpha,\beta)$ if $C_p(p,\alpha) = -D_p(p,\beta).$

Therefore, by (x) and (xi),(xii) implies:

$$\forall\,\alpha \in E, \forall\,\beta \in F, C_p(p(\overline{a}^*(\alpha),\overline{b}^*(\beta),\alpha) = -D_p(p(\overline{a}^*(\alpha),\overline{b}^*(\beta),\beta),$$

which by (xiii) contradicts (ix).

■

4. Distribution rules and collective efficiency

4.1. We have studied distribution rules in terms of their incentive compatibility properties. It may well be the case that, after paying the amounts described by the rules to the players, the budget of the agency exhibits a substantial surplus or a deficit. In terms of collective efficiency it seems necessary to introduce explicitly some budgetary consideration. For this purpose we define a new property of distribution rules: a distribution rule (F_I, F_{II}) is called a *Balanced Distribution Rule* (BDR) if and only if:

$$\forall\, \alpha \in A, \forall\, \beta \in B, F_I(a,b) + F_{II}(a,b) = C(p(a,b),a) + D(p(a,b),b).$$

In the following we show first that no DDR may be also a BDR. This in some sense (even if it could be expected) is unfortunate because of the stronger incentive property of any DDR. Then, under the restriction that the beliefs of the player should be independent of their own characteristics (conditions (1)–(2)), we show that it is always possible to find a distribution rule which is both a SDDR and a BDR. Such a rule has thus the property of being incentive compatible for each player and allows the agency to keep its budget balanced.

4.2. We begin with the negative result, namely that no DDR can be balanced.

Proposition 4 *The class of all DDR's is disjoint from the class of all BDR's.*

Proof By definition a BDR (F_I, F_{II}) is such that:

$$\forall\, a \in A, \forall\, b \in B, F_I(a,b) + F_{II}(a,b) = C(p(a,b),a) + D(p(a,b),b).$$

Hence, for any $b \in B$ and $a \in A$:

$$\begin{aligned}\frac{\partial}{\partial a}F_I(a,b) + \frac{\partial}{\partial a}F_{II}(a,b) &= \frac{\partial}{\partial a}p(a,b)[C_p(p(a,b),a) + D_p(p(a,b),b)] + C_a(p,a)\big|_{p=p(a,b)}\\ &= C_a(p,a)\big|_{p=p(a,b)} \text{ (by rule } r_1)\end{aligned}$$

Now suppose in addition that (F_I, F_{II}) is a DDR. Then we get:

$$\frac{\partial}{\partial a}F_{II}(a,b) = C_a(p,a)\big|_{p=p(a,b)}.$$

Also,

$$\begin{aligned}\frac{\partial^2}{\partial b \partial a}F_{II}(a,b) &= \frac{\partial}{\partial b}p(a,b)\left[C_{ap}(p,a)\big|_{p=p(a,b)}\right]\\ &= \frac{\partial}{\partial b}p(a,b)\left[C_{pa}(p,a)\big|_{p=p(a,b)}\right],\end{aligned}$$

which is different from zero by the remark page 7 and by condition c_5. But this is impossible since the distribution rule is a DDR. ■

For the last result we need again to restrict the admissible density functions $\mu(\cdot \mid \alpha)$ and $\nu(\cdot \mid \beta)$ by introducing the same two independence conditions (1) and (2).

Proposition 5 *If the densities μ and ν satisfy respectively (1) and (2), then there exists a distribution rule which is both a SDDR and a BDR.*

Proof First, we may define for every $a \in A$

$$f(a) \overset{\text{def}}{=} \int_B \big(C(p(a,b),a) + D\big(p(a,\beta),\beta\big)\big)\mu(\beta)(-d\beta)$$
$$g(b) \overset{\text{def}}{=} \int_A \big(C(p(\alpha,b),\alpha) + D\big(p(\alpha,b),b\big)\big)\nu(\alpha)d\alpha.$$

By an argument similar to the one of Proposition 2, it may be shown that f and g are continuously differentiable on $\mathbb{R}^n$.
Now, we can construct the required distribution rule:

$$\forall\,(a,b) \in A \times B,\; F_I(a,b) \overset{\text{def}}{=} \frac{1}{2}\Big\{C(p(a,b),a) + D(p(a,b),b) - f(a) + g(b)\Big\}$$
$$F_{II}(a,b) \overset{\text{def}}{=} \frac{1}{2}\Big\{C(p(a,b),a) + D(p(a,b),b) + f(a) - g(b)\Big\}.$$

It is immediate that (F_I, F_{II}) is a BDR. It remains to show that it is a SDDR. But, clearly

$$\int_B F_I(a,\beta)\mu(\beta)d\beta = \frac{1}{2}\int_B g(\beta)\mu(\beta)d\beta$$

which is constant in a. Similarly

$$\int_A F_{II}(\alpha,b)\nu(\alpha)d\alpha = \frac{1}{2}\int_A f(\alpha)\nu(\alpha)d\alpha$$

which is constant in b.
Hence (F_I, F_{II}) is a SDDR. ■

One easy corollary of Propositions 1 and 5 is that under assumptions (1) and (2), one may find a distribution which is both an ICDR and a BDR by letting

$$F_I(a,b) = \frac{1}{2}\big[C(p(a,b),a) + D(p(a,b),b) - f(a) + g(b)\big] - h_I(a,b) \tag{4}$$
$$F_{II}(a,b) = \frac{1}{2}\big[C(p(ab),a) + D(p(a,b),b) + f(a) - g(b)\big] - h_{II}(a,b) \tag{5}$$

where h_I and h_{II} are functions from $A \times B$ to $\mathbb{R}$ with h_I + $h_{II} \equiv 0$ and such that

$$\int_B h_I(\cdot,\beta)\nu(\beta)d\beta \text{ and } \int_A h_{II}(\alpha,\cdot)\nu(\alpha)d\alpha \text{ are constant.}$$

Conversely, any distribution rule (f_I, f_{II}) which is an ICDR and a BDR is such that there exist continuously differentiable functions h_I and h_{II} on $A \times B$, such that $h_I + h_{II} \equiv 0$, with $\int_B h_I(\cdot,\beta)\nu(\beta)d\beta$ and $\int_A h_{II}(\alpha,\cdot)I(\alpha)d\alpha$ constant and which satisfy conditions (4) and (5). We thus have a characterization of the distribution rules which are both ICDR and BDR, under our assumptions.

Among the assumptions, conditions (1) and (2) may appear quite restrictive from an informational point of view. Weaker conditions have been introduced in d'Aspremont and Gérard-Varet (1979a) to show the existence of a distribution rule with is both ICDR and BDR, when A and B are discrete finite spaces.

References

Claude d'Aspremont and Louis-André Gérard-Varet. Incentives and incomplete information. *Journal of Public economics*, 11(1):25–45, 1979a.

Claude d'Aspremont and Louis-André Gérard-Varet. On bayesian incentive compatible mechanisms. In J.J. Laffont, editor, *Aggregation and Revelation of Preferences*. North Holland, Amsterdam, 1979b.

Wendell Fleming. *Functions of several variables*. Addison Welsey Publishing Company, Inc., Reading, Mass., 1965.

Jerry R Green and Jean-Jacques Laffont. *Incentives in public decision making*. North-Holland, 1979.

Paul R Halmos. *Measure theory*. Van Nostrand Reinhold Company, New York, 1959.

John C Harsanyi. Games with incomplete information played by 'Bayesian' players. Parts I–III. *Management science*, 14:158–182, 320–334, 486–502, 1967-68.

Jean-Jacques Laffont and Eric Maskin. A differential approach to expected utility maximizing mechanisms. In J. J. Laffont, editor, *Aggregation and Revelation of Preferences*. North Holland, Amsterdam, 1979.

Henri Smets. Le principe de la compensation réciproque. un instrument économique pour la solution de certains problèmes de pollution transfrontière. O.C.D.E., Direction de l'Environnement, 1973.