European Journal of Political Economy 5 (1989) 203–217. North-Holland

STOCHASTIC INCENTIVE PROBLEMS WITH PARTIAL DYNAMIC INFORMATION AND MULTIPLE LEVELS OF HIERARCHY*

Tamer BAŞAR

University of Illinois, Urbana, IL 61801, USA

We formulate and solve a class of three-agent incentive decision problems with strict hierarchy and decentralized information. The agent at the top of the hierarchy (leader) observes a random linear combination of the decisions of the other two agents and constructs his policy based on this, as well as some static information. We show that for general concave utility functions, and under some reasonable conditions on the random variables involved, the leader has an optimal incentive policy which is linear in the partial dynamic measurement and which induces the desired behavior on the two followers.

1. Introduction

We consider, in this paper, a class of stochastic incentive decision problems with three agents, a strict hierarchy and decentralized information. The information structure is such that the two agents at the upper level of the decision hierarchy receive static as well as dynamic information, whereas the agent at the bottom of the hierarchy receives only static information on the random variable characterizing the unknown state of the environment. The dynamic information of the agent at the top of the decision hierarchy does not involve separate observations of the actions of the other two agents, but rather a single dynamic measurement which depends on both action variables. Then, the question we raise in the paper is, given that the leader has some 'ideal' point in the product decision space (one that is obtained, for example, by global maximization of his expected utility function), whether he can construct an incentive policy (using his partial dynamic information) which induces the right behavior on the followers, to lead to the desired goal. We are particularly interested in obtaining smooth incentive policies and not (discontinuous) threat policies. We will ask this question also for the case when the leader does not have any direct influence on the utility function of the agent at the bottom of the hierarchy.

The stochastic framework adopted in the paper enables us to derive some

*Research that led to this paper was supported in part by a Grant DE-FG02-88-ER-13939 from the US Department of Energy.

strong results for this class of models. In particular, we will show that for general concave utility functions, and under some reasonable conditions on the random variables involved, there exists an optimal incentive policy for the leader, linear in the partial dynamic measurement, which induces the desired behavior on the two followers. Also, there exist some additional degrees of freedom in the leader's incentive policy, which can be used to further desensitize its performance to parametric changes in the utility functions of the other agents. These results constitute extensions of some of our earlier work on the topic [such as Başar (1984), Cansever and Başar (1985)] to multiple hierarchies.

The organization of the paper is as follows: In section 2, we provide a precise mathematical formulation for the problem by concentrating on the case of scalar random variables. In section 3, we present our main result and discuss the implications of the various conditions involved. Section 4 includes an illustrative example which involves a duopolistic market with government (price) coordination and regulation. Finally, section 5 introduces some extensions to the basic model, such as vector-valued variables and multiple agents at each level, and outlines the results to be expected in each case.

2. Problem formulation

The basic ingredients of the problem we address in this paper are the following: There are three decision makers (*DM*'s), where one of them, *DM0*, referred to as 'leader', declares an incentive policy at the start of the decision process. The other two decision markers, *DM1* and *DM2*, called the 'followers', act under this incentive policy according to a given protocol. More specifically, we will assume a strict hierarchy under which *DM1* first decides on his policy (which could also be called an incentive policy in this remaining portion of the game), and announces it as enforcement on *DM2* who in turn acts optimally (under his own utility function).

Let a_i denote the action (decision) variable of *DMi*, taking values in the action (decision) space A_i, $i=0,1,2$. Let x be a random variable taking values in a set X and defined on a given probability space which is common to all three *DM*'s. This random variable denotes the 'payoff relevant' portion [à la Marschak (1963)] of the unknown state of nature. *DM1* observes the value of a (related) random variable y_1 and *DM2* observes the value of another random variable y_2, which are defined on the common probability space and take values in Y_1 and Y_2, respectively. Furthermore, *DM0* observes the value of a fourth random variable z, defined on the same probability space and taking values in a set Z; this measurement will be called the private information of the leader, since it is not shared by the other *DM*'s. It is assumed that *DM0* has also access to the observed values of y_1 and y_2, and

DM1 has access to y_2 in addition to y_1. To make the problem non-trivial, we consider only the case when all the three random variables, y_1, y_2 and z, are correlated with x.

The observation variables y_1, y_2, z introduced above all provide static information, in the sense that they do not depend on the actions of the *DM*'s. The dynamic information, which is an essential ingredient of incentive decision problems, is specified as follows for the problem under consideration. Let k_1 and k_2 be two non-zero random variables defined on the common probability space and taking values on the real line. These are also private information to the leader, and as a special case they could be taken as functions of z. They are also allowed to be correlated with y_1 and y_2, and the only restriction we impose on their selection is that when conditioned on (z, y_1, y_2) they be independent of x. In terms of these two variables, *DM0* makes the dynamic measurement

$$a = k_1 a_1 + k_2 a_2, \tag{2.1}$$

which is a (random) linear combination of the followers' actions. This partial dynamic information may, for example, correspond to the case where the leader observes the 'state' of the underlying system which evolves according to an equation linear in the followers' actions, but he cannot necessarily observe these actions separately. We also point to one implicit assumption made in writing (2.1), which is that the action spaces of the two followers are compatible so that the summation makes sense; this will indeed be the case in our formulation. Now, to complete the specification of the dynamic information, and consistent with the hierarchical mode of play, we further assume that the first follower, *DM1*, has access to the action variable of the second follower, *DM2*, and hence can base his (incentive) policy on a_2, as well as the static information (y_1, y_2). Finally, in order to convey the main ideas succinctly, we will take the decision spaces A_0, A_1, A_2, as well as the measurement spaces Y_1, Y_2, Z as copies of the real line, which we henceforth denote by R.

Admissible policies for *DM0*, *DM1* and *DM2* are g_0: $R^6 \to R$, g_1: $R^3 \to R$, and g_2: $R \to R$, respectively, with

$$a_0 = g_0(a, k_1, k_2, z, y_1, y_2) \tag{2.2a}$$

$$a_1 = g_1(a_2, y_1, y_2) \tag{2.2b}$$

$$a_2 = g_2(y_2). \tag{2.2c}$$

Furthermore, we let h_0: $R^3 \to R$ [i.e. $h_0 = h_0(z, y_1, y_2)$] and h_1: $R^2 \to R$ [i.e. $h_1 = h_1(y_1, y_2)$] denote 'static' policies for *DM0* and *DM1*, respectively. All these functions are assumed to satisfy the usual regularity conditions, discussed in Başar (1984), which are that the functions be jointly measurable in their arguments and they induce well-defined random variables on the original probability space. We let G_i denote the convex policy space where g_i belongs, $i=0,1,2$, and let H_j denote the function space where h_j belongs, $j=0,1$.

Let $u_i(x,a_0,a_1,a_2)$ denote the utility function of *DMi* ($i=0,1,2$), whose expected value, U_i, is to be maximized by him, under the three-level Stackelberg equilibrium concept [cf. Başar and Olsder (1982)], with *DM0* being the leader, *DM1* the first follower and *DM2* the second follower. The objective functional of *DMi* is the expected utility U_i, defined on the product policy space $G_0 \times G_1 \times G_2$ and given by

$$U_i(g_0,g_1,g_2)=\mathrm{E}\{u_i(x,a_0,a_1,a_2)\}, \tag{2.3}$$

where for each i a_i is related to g_i through (2.2), and $\mathrm{E}\{\cdot\}$ denotes the expectation operation over the prior statistics of the sextuple (x,z,y_1,y_2,k_1,k_2).

We now introduce a set of assumptions which make the underlying optimization problems well defined.

Assumption 2.1

(1) The policy spaces $\{G_j\}$ and the utility functions $\{u_i\}$ are defined in such a way that the expected utility (2.3) is finite for every g_j in G_j, $j=0,1,2$.
(2) For each x in R, u_0 and u_1 are concave and continuously differentiable in the triple (a_0,a_1,a_2), and are strictly concave in the pair (a_1,a_2).
(3) For each x in R, u_2 is continuously differentiable in (a_0,a_1,a_2), concave in (a_0,a_1) and strictly concave in a_2.

Now let the triple (h_0^t,h_1^t,g_2^t) provide a global maximum (i.e. a team-optimal solution) for U_0 on $H_0 \times H_1 \times G_2$. Since u_0 depends only on the random variable x, and k_1, k_2 are conditionally independent of x, nothing can be gained as far as this maximum goes by including also this private information in the information set of *DM0*; that is, the maximum would have been the same even if we had expanded H_0 so as also to include these two random variables. We also note that whenever it exists, such a maximum is unique because of the strict concavity assumption on u_0. To ensure existence of a maximum, it is possible to impose some topological conditions on $H_0 \times H_1 \times G_2$, but we will not go into such technical details here since they are only peripheral to the developments to follow. Even if such a maximum does not exist, we can view the triple (h_0^t,h_1^t,g_2^t) to represent one set of static policies, which leads to a performance, U_0^t, that is acceptable to the leader. The main result of the paper, to be presented in the next section, is that there indeed exists an incentive policy for the leader (an element of G_0) such that under the rules of the game, and under some additional conditions, the desired performance is attained.

3. Main result

Before presenting the main result alluded to above, we first introduce some new notation and terminology, and prove two auxiliary lemmas.

Let $\mathrm{E}\{(\cdot)|r\}$ denote the conditional expectation of a random variable $(\cdot)$ based on the observed value of some (other) random variable r. Let $\nabla_p f(.,p,.)$ denote the first-order partial derivative of a function f with respect to the indicated argument, p. Let U_0^t denote an acceptable level of performance for the leader, achieved by the triple (h_0^t, h_1^t, g_2^t), as introduced in section 2. For short, let

$$a_0^t := h_0^t(z, y_1, y_2);\ a_1^t := h_1^t(y_1, y_2);\ a_2^t := g_2^t(y_2), \quad \text{and define} \tag{3.1}$$

$$a^t := k_1 a_1^t + k_2 a_2^t. \tag{3.2}$$

Introduce the functions F_0: $R^3 \to R$, F_1: $R^2 \to R$, F_0: $R \to R$ defined by

$$F_0(z, y_1, y_2) := \mathrm{E}\{\nabla_{a_0} u_1(x, a_0^t, a_1^t, a_2^t) | z, y_1, y_2\}, \tag{3.3a}$$

$$F_1(y_1, y_2) := \mathrm{E}\{\nabla_{a_1} u_1(x, a_0^t, a_1^t, a_2^t) | y_1, y_2\}, \tag{3.3b}$$

$$F_2(y_2) := \mathrm{E}\{\nabla_{a_2} u_1(x, a_0^t, a_1^t, a_2^t) | y_2\}. \tag{3.3c}$$

Finally, let $b := (z, y_1, y_2, k_1, k_2)$, and introduce the pair of equations

$$\mathrm{E}\{Q_0 F_0(z, y_1, y_2) k_1 | y_1, y_2\} = F_1(y_1, y_2), \tag{3.4a}$$

$$\mathrm{E}\{Q_0 F_0(z, y_1, y_2) k_2 | y_2\} = F_2(y_2), \tag{3.4b}$$

in terms of a b-measurable random variable Q_0.

Condition 3.1. There exists at least one b-measurable random variable Q_0, mutually satisfying (3.4).

Remark 3.1. First note that by taking the expected value of both sides of (3.4a) over y_1, and conditioned on y_2, one arrives at

$$\mathrm{E}\{Q_0 F_0(z, y_1, y_2) k_1 | y_2\} = \mathrm{E}\{F_1(y_1, y_2) | y_2\},$$

which is a necessary condition for (3.4a). Further note that in this equation [as well as in (3.4a)] k_1 can be replaced, without any loss of generality, by $\mathrm{E}\{k_1|z, y_1, y_2\} =: \bar{k}_1$; likewise, in (3.4b) k_2 can be replaced by its conditional expectation with respect to the same set of measurements, which we denote by $\bar{k}_2$. From these observations, it follows that a necessary condition for existence of at least one solution to (3.4) is that the two random variables $\bar{k}_1$ and $\bar{k}_2$ be linearly independent (that is, there is no constant e such that $\bar{k}_1 = e\bar{k}_2$), unless the random variables $F_2(y_2)$ and $\mathrm{E}\{F_1(y_1, y_2)|y_2\}$ are linearly dependent.

Lemma 3.1. Let Condition 3.1 be satisfied, and Q_0 be one such solution. Then there exists a policy g_0^* for the leader,

$$g_0^*(a, b) = h_0^t(z, y_1, y_2) - Q_0(a - a^t), \tag{3.5}$$

under which the unique solution to the problem of maximizing $U_1(g_0^*, h_1, g_2)$ over $H_1 \times G_2$ is (h_1^t, g_2^t).

Proof. Note that the optimization problem

$$\max_{h_1, g_2} U_1(g_0^*, h_1, g_2)$$

is in fact a two-agent stochastic team problem with a static information structure. The first-order necessary conditions for an optimum (which are obtained by first holding g_2 fixed and taking variation with respect to h_1, and then reversing the roles of the two variables) are:

$$0 = (d/da_1)\mathrm{E}\{u_1(x, a_0^t - Q_0(k_1 a_1 + k_2 a_2 - a^t), a_1, a_2)|y_1, y_2\} \tag{*}$$

$$0 = (d/da_2)\mathrm{E}\{u_2(x, a_0^t - Q_0(k_1 a_1 + k_2 a_2 - a^t), a_1, a_2)|y_2\}. \tag{**}$$

In writing the above we have utilized the fact that any functional optimization problem

$\max_{a=h(y)} \mathrm{E}\{f(x, a)\}$ can be written as

$\mathrm{E}\{\max_a \mathrm{E}\{f(x, a)|y\}\}$.

Now, taking the total derivative in (*) and (**) in terms of partial derivatives after interchanging that operation with conditional expectation (this interchange is valid under our smoothness assumptions), we arrive at the following equivalent equations (where we suppress the arguments):

$$0=\mathrm{E}\{\nabla_{a_1}u_1-\nabla_{a_0}u_1\cdot Q_0k_1|y_1,y_2\} \tag{*'}$$

$$0=\mathrm{E}\{\nabla_{a_2}u_1-\nabla_{a_0}u_1\cdot Q_0k_2|y_2\}. \tag{**'}$$

If we choose $a_1=h_1^t(y_1,y_2)$, $a_2=g_2^t(y_2)$ in the above, it follows that whenever Q_0 satisfies (3.4), (*′) and (**′) become identities. Thus, (h_1^t,g_2^t) constitutes a pair of stationary policies for the stochastic team problem. By Theorem 1 (p. 863) of Radner (1962) they are also Bayes (i.e. team-optimal). Furthermore, they are unique as Bayes rules because of strict concavity of u_1 in all three action variables, for every value of x (cf. Assumption 2.1), in addition to the fact that the chosen policy for *DM0* is linear which in turn is linear in a_1 and a_2. We note that the 'local finiteness' condition of Theorem 1 of Radner (1962) is satisfied here because of our Assumption 2.1.

In words, what Lemma 3.1 says is that if *DM2* had the same utility function as *DM1* (though still operating under different information), then the incentive policy (3.5) by the leader would induce him (i.e. *DM2*) and *DM1* to act jointly in such a way so as to maximize *DM0*'s expected utility. In our formulation, however, the second follower has a (generally) different utility function, and hence pursues a different set of goals than *DM1*. The question then is whether under the assumed mode of play (i.e., strict hierarchy), the leader would still be able to enforce his most favorable solution. The answer is in the affirmative, as elucidated in the sequel.

Working towards our main result, we first observe that if the leader announces the policy (3.5), and Condition 3.1 is satisfied, then the best performance the first follower (*DM1*) can expect to achieve is $U_1(g_0^*,h_1^t,g_2^t)$, which requires full cooperation by the second follower. Since *DM2* is in reality maximizing a different objective functional, the cooperation is not there, unless *DM1* can enforce it by using his dynamic information. This will indeed be possible if there exists a policy for *DM1*, say g_1^*, in G_1, under which the optimal response of *DM2*, computed through the maximization problem

$$\max_{g_2} U_2(g_0^*,g_1^*,g_2) \tag{3.6a}$$

is uniquely given by g_2^t, and furthermore that

$$g_1^*(g_2^t(y_2),y_1,y_2)\equiv h_1^t(y_1,y_2). \tag{3.6b}$$

Lemma 3.2 below says that such an incentive policy exists for *DM1* under some fairly reasonable conditions. Towards this end, first we introduce three functions: $f_0: R^3 \to R, f_1: R^2 \to R, f_2: R \to R$, defined by

$$f_0(z, y_1, y_2) := \mathrm{E}\{\nabla_{a_0} u_2(x, a_0^t, a_1^t, a_2^t) | z, y_1, y_2\} \tag{3.7a}$$

$$f_1(y_1, y_2) := \mathrm{E}\{\nabla_{a_1} u_2(x, a_0^t, a_1^t, a_2^t) | y_1, y_2\} \tag{3.7b}$$

$$f_2(y_2) := \mathrm{E}\{\nabla_{a_2} u_2(x, a_0^t, a_1^t, a_2^t) | y_2\}. \tag{3.7c}$$

Second, with $c := (y_1, y_2)$, we introduce a c-measurable random variable Q_1, which satisfies the linear equation:

$$\mathrm{E}\{Q_0 f_0(z, y_1, y_2)[k_2 - k_1 Q_1] | y_2\} + \mathrm{E}\{Q_1 f_1(y_1, y_2) | y_2\} = f_2(y_2), \tag{3.8}$$

where Q_0 is obtained from (3.4).

Condition 3.2. The function $f(c)$ defined by

$$f(c) := f_1(c) - \mathrm{E}\{Q_0 f_0(z, y_1, y_2) k_1 | c\} \tag{3.9}$$

is non-singular for every $c := (y_1, y_2)$ in R^2.

Lemma 3.2. Let Condition 3.2 be satisfied. Then,

(*i*) *The linear equation* (*3.8*) *admits at least one solution* Q_1.
(*ii*) *There exists a policy* g_1^* *for DM 1,*

$$g_1^*(a_2, c) = h_1^t(y_1, y_2) - Q_1(a_2 - a_2^t) \tag{3.10}$$

where Q_1 *is any c-measurable solution of* (*3.8*), *under which the maximization problem* (*3.6a*) *admits the unique solution* g_2^t, *and furthermore the side condition* (*3.6b*) *is satisfied.*

Proof. The first part of the Lemma follows from Proposition 2 (p. 205) of Başar (1984). To prove the second part, we take the variation of $U_2(g_0^*, g_1^*, g_2)$ with respect to g_2, which is equivalent to taking the total derivative of $U(a_2, y_2) := \mathrm{E}\{u_2(x, g_0^*(a, b), g_1^*(a_2, c), a_2) | y_2\}$ with respect to a_2. With g_0^* and g_1^* taken as in (3.5) and (3.8), respectively, where Q_0 and Q_1 satisfy (3.4) and (3.8), it is not difficult to see that the derivative is identically zero when $a_2 = g_2^t(y_2)$. Hence g_2^t is a stationary point of $U_2(g_0^*, g_1^*, g_2)$. Since u_2 was taken to be concave in (a_0, a_1) and strictly concave in a_2, and since g_1^* and in turn g_0^* are linear in a_2, it follows that $U(a_2, y_2)$ is strictly concave in a_2 for every y_2, and hence its stationary point is the unique maximum for each y_2. Finally, the side condition (3.6b) is clearly satisfied by the policy (3.10), as can be seen by inspection.

Remark 3.2. Using the function f defined by (3.9), eq. (3.8) can be rewritten in the simpler form

$$\mathrm{E}\{f(c)Q_1|y_2\}=f_2(y_2)-\mathrm{E}\{Q_0f_0(z,y_1,y_2)k_2|y_2\}=:m(y_2). \tag{3.8$'$}$$

Let w be any c-measurable random variable such that $\mathrm{E}\{wf(c)|y_2\}\neq 0$ for every y_2 in R. Then the choice $Q_1=wm(y_2)/\mathrm{E}\{wf(c)|y_2\}$ is a solution to (3.8′), as can be verified by direct substitution. Since there will in general be many w's meeting the non-singularity condition, the solution to (3.8) will in general be non-unique. This provides us with additional degrees of freedom in the final choice of Q_1, and leaves room for the introduction of some additional criteria involving issues like 'minimum sensitivity' and 'robustness', as in Cansever and Başar (1985). For example, the optimal choice could desensitize the performance to parametric changes in the utility function of *DM2*.

We are now in a position to present the main theorem of this section, by combining the results of Lemmas 3.1 and 3.2.

Theorem 3.1. Let U_0^t be an acceptable performance level by the leader, which is achieved by some triple (h_0^t,h_1^t,g_2^t) in $H_0\times H_1\times G_2$. In terms of this triple, let Conditions 3.1 and 3.2 be satisfied. Then, under the hierarchical mode of play, the incentive policy g_0^* for the leader, given by (3.5), forces the two followers to a cooperative solution, leading to achievement of the ideal performance U_0^t.

Proof. The result follows directly from Lemmas 3.1 and 3.2, and the discussion preceding Lemma 3.2.

We conclude this section by making a number of important observations on the general solution presented in Theorem 3.1.

(1) The Theorem covers the following two extreme cases: (i) the leader does not have any direct influence on the utility function of *DM2* (i.e. $\nabla_{a_0}u_2\equiv 0$) and (ii) *DM1* does not have any direct influence on u_2 (i.e. $\nabla_{a_1}u_2\equiv 0$). In either case, the leader is able to force both followers to a Pareto-optimal solution to his utmost advantage. In the former case, this implies that the inducement provided by the leader's optimal incentive policy propagates down the hierarchy and its effect is felt by the second follower (who has no direct link to *DM0*) through the first follower's policy choice.

(2) The coefficients k_1 and k_2 in the leader's dynamic measurement play an

important role in the final result and the corresponding existence conditions. As hinted earlier in Remark 3.2, since k_1 and k_2 influence eqs. (3.4) and (3.8) only through $\bar{k}_1 := \mathrm{E}\{k_1|z, y_1, y_2\}$ and $\bar{k}_2 := \mathrm{E}\{k_2|z, y_1, y_2\}$, we can take them to be (z, y_1, y_2)-measurable, without any loss of generality. For Condition 3.2 to hold, it will generally be necessary for $\bar{k}_1$ and $\bar{k}_2$ to be linearly independent, which rules out the possibility of choosing them as deterministic parameters. Hence, if k_1 and k_2 are viewed as additional design parameters, to be chosen by *DM0* based on his static information, then they should be 'stochastically rich'. This makes the consequences of the leader's policy decision (based primarily on his private information) unpredictable by the two followers, with this uncertainty forcing them to a full cooperation. Note that with a deterministic (purely predictable) information pattern, such a cooperative solution can never be enforced. It is perhaps ironical that the less the two followers know of the consequences of their actions (such as the reward structure) the better it is for the leader (in terms of enforcement of the cooperative solution), provided of course that the uncertainty is stochastic with a common distribution known to all parties.

(3) As was mentioned earlier, the ideal solution (h_0^t, h_1^t, g_2^t) does not have to be chosen as a utility maximizing solution (to the leader). It may also be construed as a 'compromise' solution, which takes into account the welfare of all the *DM*'s involved. Provided that the two conditions are satisfied, such a compromise solution can indeed be enforced using the smooth incentive policy g_0^* given by (3.5).

4. An example

We provide, in this section, a simple example to illustrate the theoretical results of the previous section. We consider a duopolistic market with a hierarchical structure involving, say, a state enterprise (*DM1*) and a private firm (*DM2*), producing identical products. A third party (the government – *DM0*) regulates the production level by controlling the demand (thereby the price) for the product. This may occur, for example, if the government is the sole buyer of the product.

There are two possible states of the (economic) environment, which affect desired levels of production as well as the desired price level (by the government). In accordance with the earlier notational convention, we denote this discrete variable by x, and the two possible states by x_1 and x_2. We assume that x_1 and x_2 are equally likely (that is each occurs with probability 0.5), and they are observable by the government. The other two agents, *DM1* and *DM2*, however, do not know the true occurrence of x (particularly at the time they decide on their production levels), but they know that it takes only the given two values, and with equal probability. In addition to this common knowledge, *DM1* can make some inference of the true state, but with some error. More precisely, *DM1* makes a correct assessment of the true state (x_1 or x_2) each with probability 0.6, and errs in each case again with equal probability 0.4. Mathematically speaking, *DM1* observes the true value of a discrete variable y which takes two values, y_1 and y_2, with the conditional probability of the event $\{y=y_i\}$ given that $x=x_j$ being 0.6 for $i=j$ and 0.4 for i different from j $(i,j=1,2)$. Note that, with this construction, the random variable y takes the given two values with equal probability 0.5. In the general formulation of section 2, y corresponds to y_1, x here corresponds to both x and z (that is, they are identical), and the variable y_2 is vacuous.

The action variables a_0, a_1, a_2 are, respectively, the price level set by the government and the production levels of *DM1* and *DM2*. The government makes its price policy decision based also on some observable a that involves a random linear combination of the production levels of the two firms, as in (2.1). Here, k_1 and k_2 are two random variables which are not directly observable (particularly at the time production decisions are made), but they could be correlated with the two random variables x and y. The eq. (2.1) could admit the interpretation that *DM0* observes the output of a process that uses both a_1 and a_2 as inputs, with the production process not being totally deterministic. He is, however, allowed to know their realized values at the time the price policy is implemented.

Now, let us assume that *DM0* (the government) determines some desirable price and production levels (which could be the result of some welfare maximization problem). These will have to be compatible with the information available to the decision makers, since the objective is to enforce them (smoothly – via incentives) on the two firms (the followers). Let us denote these, as before, by a_0^t, a_1^t and a_2^t, respectively. Their dependence on the various variables introduced above would be [following also from (3.1)]:

$$a_0^t := h_0^t(x, y);\; a_1^t := h_1^t(y);\; a_2^t := \text{a constant}.$$

Let us introduce, as before, $b := (x, y, k_1, k_2)$. Then, the theory of section 3 says that, provided that the utility functions of *DM1* and *DM2* have the right concavity properties, and some additional conditions specified there are satisfied, there will exist an optimal price policy for the government, in the form

$$g_0^*(a,b) = h_0^t(x,y) - Q_0(a - a^t), \tag{4.1}$$

where Q_0 is some b-measurable random variable [satisfying (3.4)]. Such a policy will force the two firms to the desired production levels a_1^t, a_2^t, under the stipulation that there is a hierarchy between these two firms, and *DM1* can enforce a 'smooth' production policy on *DM2*. This policy will also be induced by (4.1), and will be in the form [from (3.10)]:

$$g_1^*(a_2, y) = h_1^t(y) - Q_1(a_2 - a_2^t), \tag{4.2}$$

where Q_1 is some y-measurable random variable, satisfying (3.8). Note that this policy shows explicit dependence on the production level of *DM2*.

We now construct these policies for a specific set of utility functions for *DM1* and *DM2*, and show that all the conditions of Lemma 3.1 and Lemma 3.2 are satisfied, even if some of the concavity conditions are relaxed. Toward this end, we first take the utility function of *DM2* as the standard duopolistic profit function with a linear cost of production [see, for example, Friedman (1983)]:

$$u_2 = a_0 a_2 - (1/2)a_2. \tag{4.3}$$

It is of interest to note that the utility function of *DM2*, as given above, does not explicitly depend on the action variable of *DM1*. For *DM1*, on the other hand, we adopt the utility function

$$u_1 = a_0 a_1 - (1/2)a_1 + (1/10)u_2, \tag{4.4}$$

where the first two terms are as in (4.4), and the third term is included so as to provide some incentive for *DM1* to also take into account *DM2*'s profit maximization while developing his production policy. This term could be interpreted as a subsidy to *DM1* through taxation of *DM2*. To make the problem symmetric, a similar subsidy term (based on a fraction of *DM1*'s net profit) could have been included in (4.3), but we will not pursue that symmetric case here.

Toward obtaining some numerical results, we take the desired price level a_0^t to be 1 when $x = x_1$ and to be 2 when $x = x_2$. Furthermore, we take a_1^t to be 1 when $y = y_1$ and to be 2 when $y = y_2$, and take a_2^t to be equal to 1. Finally, we choose k_2 to be 1 (that is, a degenerate random variable), and k_1 to be x-measurable and taking the two values 1 and 2 when x is x_1 and x_2 respectively. Let the support set of the random variable Q_0 be $\{Q_{011}, Q_{012}, Q_{021}, Q_{022}\}$, where Q_{0ij} is the value Q_0 takes when $x = x_i$ and $y = y_j$, for $i, j = 1, 2$. Then, the pair of eqs. (3.4) can be rewritten in terms of the Q_{0ij}'s as three equations:

$$6.6Q_{011} + 8.8Q_{021} = 9$$

$$8.4Q_{012} + 25.2Q_{022} = 11$$

$$6Q_{011} + 4Q_{021} + 8Q_{012} + 12Q_{022} = 2.$$

This set of equations admits multiple solutions, any one of which would be an acceptable choice. To produce a candidate, we force the symmetry condition $Q_{011}=Q_{022}$, under which the unique solution is

$$Q_{011}=Q_{022}=1.3963443; \quad Q_{012}=-2.879508; \quad Q_{021}=-0.024531. \tag{4.5a}$$

Using this in (3.8), we obtain the constant solution

$$Q_1=1.1073114. \tag{4.5b}$$

Of course, (3.8) would admit other solutions which would be functions of the variable y, but the one above should be preferable since it is a constant.

Hence the example of this section admits multiple solutions, all in the form (4.1)–(4.2), for *DM0* and *DM1*, with one set of coefficient values given by (4.5). In simpler form, the government's optimal demand curve is given by the linear relation

$$a_0 \equiv g_0^*(a,b)=a_0^t-Q_0(a-k_1a_1^t-1), \tag{4.6}$$

where Q_0 is given by (4.5), and a_0^t, a_1^t are the desirable levels for price and production (for *DM1*) as given earlier. Faced with this price policy, and taking the existing hierarchy into consideration, an optimal decision for *DM1* becomes the production policy

$$a_1 \equiv g_1^*(a_2,y)=a_1^t-1.1073(a_2-1). \tag{4.7}$$

Under (4.6) and (4.7), the expected profit of *DM2* is maximized uniquely by the production level $a_2^t=1$, and hence the goals set at the beginning have all been met, with the inducement provided by the 'smooth' policy (4.6).

We conclude this section by making a number of useful observations.

(1) The first is of a technical nature and concerns Assumption 2.1. As we have hinted earlier, the example treated above does not meet the joint concavity requirements set there; however, the utility functions of both *DM1* and *DM2* can be made strictly concave by an appropriate choice of the leader's policy. In that case (and particularly in the context of this example) the proofs (and thereby the statements) of Lemmas 3.1 and 3.2 remain still valid, which shows that the main result of this paper holds under even less stringent conditions.

(2) The second observation relates to the fact that even though the second follower's utility function does not show explicit dependence on the first

follower's action variable, this dependence could be brought in quite naturally through an incentive policy, as we have seen in the duopoly game above.

(3) The third and final comment is that the above duopolistic framework can be extended to oligopolistic situations without much difficulty. There the single private firm at the bottom of the hierarchy would be replaced by several oligopolistic firms [in the spirit of Sertel (1988)], each of whom would have a profit function of the type (4.3), and would play a Nash game among themselves. The variable a_2 would then be the aggregate production level of the private sector, and (3.8) would be replaced by as many similar equations as the number of firms. Even though this extension is not difficult conceptually, it does require the introduction of additional excessive notation and some new terminology for the general case, which is the reason why we have avoided doing it here.

5. Some extensions

The analyses and results of this paper could be extended in several directions. The first such extension would be to the case of vector-valued action and random variables, which would cover scenarios where the agents control more than one instrument variable. We have avoided doing this here simply not to bury the basic message of section 3 in the notational complexity which would be needed in that case. We should note, however, that the gist of the results here carry over to the vector case, with the leader now having some additional degrees of freedom in the choice of the best incentive policy.

A second extension would be to problems where there are more than three levels of hierarchy. Again the analysis of this paper could be carried over to such problems, by repeated application of the ideas used in the derivation of Lemmas 3.1 and 3.2. One can show that the indirect influence of the leader on the utility maximizing behavior of the following agents is still present, provided that some conditions similar to those given in section 3 are satisfied. One could also allow for more than one agent to act at each level, either as a team or as Nash followers, with inducement towards a cooperative behavior again provided by a smooth incentive policy of the leader. To obtain some explicit results in this case, our earlier work on the two-level problem, reported in Cansever and Başar (1985), could provide some of the necessary tools.

One question that could be raised at this point is whether the results presented here would be valid if the leader had obtained imperfect (noisy) information on the followers' actions (for example, if he had not known the realized values of the two random variables k_1 and k_2 in our formulation). Another such imperfection would arise if the leader did not have access to the measurements made by the followers. In such cases, the kind of strong results obtained here would no longer exist, and one could at best hope to obtain asymptotic results of the type developed by Radner (1981, 1985). This will then require a repeated game formulation, where the leader's policy also includes a 'learning' component. Derivation of optimal 'review strategies' for such repeated games when there are multiple levels of hierarchy is a challenge for the future.

References

Başar, T., 1984, Affine incentive schemes for stochastic systems with dynamic information, SIAM J. Control and Optimization 22, no. 2, 199–210.

Başar, T. and G.J. Olsder, 1982, Dynamic noncooperative game theory (Academic Press, London/New York).

Cansever, D.H. and T. Başar, 1985, On stochastic incentive control problems with partial dynamic information, Systems and Control Letters 6, no. 1, 69–75.

Friedman, J., 1983, Oligopoly theory (Cambridge University Press, Cambridge).

Marschak, J., 1963, The payoff-relevant description of states and acts, Econometrica 31, 719–725.

Radner, R., 1962, Team decision problems, Annals of Mathematical Statistics 33, 857–881.

Radner, R., 1981, Monitoring cooperative agreements in a repeated principal–agent relationship, Econometrica 49, 1127–1148.

Radner, R., 1985, Repeated principal–agent games with discounting, Econometrica 53, no. 5, 1173–1198.

Sertel, M.R., 1988, Regulation by participation, Journal of Economics 48, no. 2, 111–134.